

NASA/CR-2001-211269  
ICASE Report No. 2001-40



# **Eigensolution Analysis of the Discontinuous Galerkin Method with Non-uniform Grids, Part I: One Space Dimension**

*Fang Q. Hu*  
*Old Dominion University, Norfolk, Virginia*

*Harold L. Atkins*  
*NASA Langley Research Center, Hampton, Virginia*

*ICASE*  
*NASA Langley Research Center*  
*Hampton, Virginia*

*Operated by Universities Space Research Association*



National Aeronautics and  
Space Administration

Langley Research Center  
Hampton, Virginia 23681-2199

Prepared for Langley Research Center  
under Contract NAS1-97046

December 2001

Report Documentation Page		
<b>Report Date</b> 00DEC2001	<b>Report Type</b> N/A	<b>Dates Covered (from... to)</b> -
<b>Title and Subtitle</b> Eigensolution Analysis of the Discontinuous Galerkin Method with Non-uniform Grids, Part I: One Space Dimension	<b>Contract Number</b>	
	<b>Grant Number</b>	
	<b>Program Element Number</b>	
<b>Author(s)</b> Fang Q. Hu, Harold L. Atkins	<b>Project Number</b>	
	<b>Task Number</b>	
	<b>Work Unit Number</b>	
<b>Performing Organization Name(s) and Address(es)</b> National Aeronautics and Space Administration Langley Research Center Hampton, Virginia 23681-2199	<b>Performing Organization Report Number</b>	
<b>Sponsoring/Monitoring Agency Name(s) and Address(es)</b>	<b>Sponsor/Monitor's Acronym(s)</b>	
	<b>Sponsor/Monitor's Report Number(s)</b>	
<b>Distribution/Availability Statement</b> Approved for public release, distribution unlimited		
<b>Supplementary Notes</b> ICASE Report No. 2001-40		
<b>Abstract</b> <p>Abstract. We present a detailed study of spatially propagating waves in a discontinuous Galerkin scheme applied to a system of linear hyperbolic equations. We start with an eigensolution analysis of the semi-discrete system in one space dimension with uniform grids. It is found that, for any given order of the basis functions, there are at most two spatially propagating numerical wave modes for each physical wave of the Partial Differential Equations (PDE). One of the modes can accurately represent the physical wave of the PDE and the other is spurious. The directions of propagation of these two numerical modes are opposite, and, in most practical cases, the spurious mode has a large damping rate. Furthermore, when an exact characteristics split flux formula is used, the spurious mode becomes non-existent. For the physically accurate mode, it is shown analytically that the numerical dispersion relation is accurate to order <math>2p + 2</math> where <math>p</math> is the highest order of the basis polynomials. The results of eigensolution analysis are then utilized to study the effects of a grid discontinuity, caused by an abrupt change in grid size, on the numerical solutions at either side of the interface. It is shown that, due to mode decoupling, numerical reflections at grid discontinuity, when they occur, are always in the form of the spurious non-physical mode. Closed form numerical reflection and transmission coefficients are given and analyzed. Numerical examples that illustrate the analytical findings of the paper are also presented.</p>		
<b>Subject Terms</b>		

<b>Report Classification</b> unclassified	<b>Classification of this page</b> unclassified
<b>Classification of Abstract</b> unclassified	<b>Limitation of Abstract</b> SAR
<b>Number of Pages</b> 31	

# EIGENSOLUTION ANALYSIS OF THE DISCONTINUOUS GALERKIN METHOD WITH NON-UNIFORM GRIDS, PART I: ONE SPACE DIMENSION

FANG Q. HU\* AND HAROLD L. ATKINS†

**Abstract.** We present a detailed study of spatially propagating waves in a discontinuous Galerkin scheme applied to a system of linear hyperbolic equations. We start with an eigensolution analysis of the semi-discrete system in one space dimension with uniform grids. It is found that, for any given order of the basis functions, there are at most two spatially propagating numerical wave modes for each physical wave of the Partial Differential Equations (PDE). One of the modes can accurately represent the physical wave of the PDE and the other is spurious. The directions of propagation of these two numerical modes are opposite, and, in most practical cases, the spurious mode has a large damping rate. Furthermore, when an exact characteristics split flux formula is used, the spurious mode becomes non-existent. For the physically accurate mode, it is shown analytically that the numerical dispersion relation is accurate to order  $2p + 2$  where  $p$  is the highest order of the basis polynomials. The results of eigensolution analysis are then utilized to study the effects of a grid discontinuity, caused by an abrupt change in grid size, on the numerical solutions at either side of the interface. It is shown that, due to “mode decoupling”, numerical reflections at grid discontinuity, when they occur, are always in the form of the spurious non-physical mode. Closed form numerical reflection and transmission coefficients are given and analyzed. Numerical examples that illustrate the analytical findings of the paper are also presented.

**Key words.** finite element methods, unstructured grids, wave propagation, acoustics

**Subject classification.** Numerical Analysis

**1. Introduction.** Discontinuous Galerkin Method (DGM) is a finite element method that allows a discontinuity of the numerical solution at element interfaces. It has been developed very rapidly in the past few years and has been applied to many fields of practical importance, such as computational fluid dynamics, aeroacoustics, and electromagnetics (e.g., [2,3,7,22]). A recent review of DGM can be found in [5] with an extensive list of references.

It is well known that for a discontinuous Galerkin scheme employing basis polynomials up to order  $p$ , the rate of convergence is  $h^{p+1/2}$  in general and  $h^{p+1}$  in some special cases ([13,12,17,19]), where  $h$  is a measure for the size of elements. Occurrences of super-convergence in DGM has been reported in the literature and some are reviewed in [5]. For examples, Biswas, Devine and Flaherty [4] and Adjerid, Aiffa and Flaherty [1] showed super-convergence on Gauss-Radau points. Lowrie, Roe and van Leer reported numerical results of order  $2p + 1$  convergence in [14]. Most recently, Cockburn, Lusk, Shu and Süli [6] showed the possibility of obtaining  $2p + 1$  convergence by a suitable post-processing of the numerical solution.

In contrast to the many studies on convergence rates, there have been relatively fewer works on the wave propagation properties of DGM. In [12], Johnson and Pitkäranta included a Fourier analysis of DGM for the case of  $p = 1$  and showed that the eigenvalue of the “amplification matrix” ( $\hat{\mathbf{E}}(\theta)$ ) is accurate to order 4 (local error). In [15], Lowrie performed a Fourier analysis of a space-time discontinuous Galerkin

---

\*Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529. Supported by the National Aeronautics and Space Administration under NASA Contract No. NAS1-97046 while the first author was in residence at ICASE, NASA Langley Research Center, Hampton, VA 23681, and by NASA Langley Grant NAG-1-01044.

†NASA Langley Research Center, Hampton, VA 23681.

scheme, up to  $p = 3$ , for a one-dimensional scalar advection equation and showed that the eigenvalue is accurate to order  $2p + 2$  (locally) which results in a global order  $2p + 1$  decay of the evolution component of the numerical error. In [9], Hu, Hussaini and Rasetarinera studied numerical dissipation and dispersion errors of DGM for one- and two-dimensional wave equations. They also analyzed anisotropic errors of wave propagation in triangular and quadrilateral elements. In a recent work by Rasetarinera, Hussaini and Hu [18], it was further demonstrated numerically that dissipation errors of DGM decay at order  $2p + 2$  (locally) when the exact characteristics splitting flux formula is used. Another study of Fourier analysis was carried out in [20] by Sherwin which gave exact expressions of the numerical frequency analytically up to  $p = 3$  and numerically for  $p = 10$ . It is also interesting to note here some related works in *continuous* Galerkin methods. For instance, a case of super-convergence in phase error in continuous Galerkin methods has been shown, numerically, by Thompson and Pinsky[21] and, theoretically, by Ihlenburg and Babuska[11]. It was found that when basis polynomials of order  $p$  are used the phase error converges (locally) at  $h^{2p+1}$ [11]. We point out that this is one order less than that for DGM as we will show in this paper.

The present work has been motivated primarily by the need to understand wave propagation in DGM with non-uniform elements (grids). As a first step toward such a goal, we study wave propagation through an interface with an abrupt change in grid size in one space dimension. We will first carry out an analysis on spatially propagating waves, referred to as the eigensolutions, of the semi-discrete system in uniform grids. Then the results of such an analysis will be applied to study wave reflection and transmission by expressing the numerical solution on either side of the interface in eigensolutions. The reflection and transmission coefficients are then found by deriving proper coupling conditions at the interface. As we will see, numerical reflection at a grid discontinuity is dependent on the flux formula employed in the implementation of DGM. Two commonly used flux schemes are considered in this paper, namely, the characteristics-based flux and Lax-Friedrich flux formulas. These two schemes will be analyzed in a unified way by introducing an upwind factor.

We point out that a major difference between the present and previous works in wave analysis for DGM is that in the present work we study spatial waves where the temporal frequency is specified and the corresponding wavenumber is sought as eigenvalues, while in the previous studies the wavenumber was specified and the frequency was found as eigenvalues. The present approach is necessary because, with an introduction of grid discontinuity, numerical wave number is not constant across the interface of a grid change. The use of spatial waves also turn out to be advantageous in that the eigenvalue problem is greatly simplified and reduced. As a result, the numerical dispersion relation is governed by a quadratic equation that can be solved analytically for any order of the basis polynomials. Specifically, in a uniform grid, it is found that there are at most two spatially propagating numerical wave modes for each physical wave of the PDE. One of the numerical wave modes can accurately approximate the physical wave and the other is a highly irregular spurious mode. They will be referred to as the *physical* and *spurious numerical waves* respectively in this paper. For the physically accurate mode, it will be shown that the numerical dispersion relation is accurate to  $(kh)^{2p+2}$  locally where  $k$  is the wave number, which confirms those previous works mentioned earlier[12][14][18]. In fact, we will show that dispersion error is of order  $2p+3$  while the dissipation error is of order  $2p+2$ . For the spurious mode, it is found that it propagates in the opposite direction of the physical mode and becomes non-existent when the exact characteristics-based flux formula is used. Following the analysis of waves in uniform grids, the effect of a grid change on either side of the interface is studied. It is found that waves associated with different physical eigenvectors are decoupled and numerical reflections are always in the form of the spurious numerical wave and are highly damped.

The rest of the paper is organized as follows. In section 2, we describe the discretization process and the associated flux formulas. In section 3, the eigenvalue problem for spatially propagating waves in a uniform grid is formulated. In section 4, numerical dispersion relation and its accuracy are analyzed and discussed. Wave propagation through a grid discontinuity is studied in section 5. And numerical examples are presented in section 6. Section 7 has our conclusions.

**2. Formulations of discretization and numerical flux.** Consider the discontinuous Galerkin method for a system of hyperbolic equations in one-dimensional space:

$$(2.1) \quad \frac{\partial \mathbf{u}}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{u})}{\partial x} = 0$$

where  $\mathbf{u}$  is a vector of dimension  $N$  and  $\mathbf{f}$  is the flux vector. We will only consider linear cases in our analysis and assume that

$$(2.2) \quad \mathbf{f}(\mathbf{u}) = \mathbf{A}\mathbf{u}$$

where  $\mathbf{A}$  is a constant  $N \times N$  matrix. We assume that  $\mathbf{A}$  has  $N$  real eigenvalues, denoted by  $a_j$  for  $j = 1, 2, \dots, N$  and the eigenvectors of  $\mathbf{A}$ , denoted by  $\mathbf{e}_j$ , form a complete basis in  $N$ -dimensional space. Throughout this paper, unless specified otherwise, lower case bold face letters will stand for column vectors and upper case bold face letters stand for matrices.

In a discretization of (2.1) using the discontinuous Galerkin method, the spatial domain is partitioned into elements,  $E_n = [x_{n-1}, x_n]$ , where  $n$  is the element index. In each element, the numerical solution, denoted by  $\mathbf{u}_h^n(x, t)$ , is expressed as

$$(2.3) \quad \mathbf{u}_h^n(x, t) = \sum_{\ell=0}^p \mathbf{c}_\ell^n(t) p_\ell^n(x),$$

where  $\{p_\ell^n(x), \ell = 0, 1, \dots, p\}$  is the set of basis polynomials for element  $E_n$ . Here  $p$ , without superscript or subscript, denotes the highest order of polynomials in the chosen basis and  $\mathbf{c}_\ell^n(t)$  is the expansion coefficient. In a weak formulation for (2.1), we require that

$$(2.4) \quad \int_{x_{n-1}}^{x_n} \left( \frac{\partial \mathbf{u}_h^n}{\partial t} + \frac{\partial \mathbf{f}}{\partial x} \right) p_{\ell'}^n(x) dx = 0$$

for  $\ell' = 0, 1, \dots, p$ . By a use of integration by parts, the above is re-written as follows,

$$(2.5) \quad \int_{x_{n-1}}^{x_n} \frac{\partial \mathbf{u}_h^n}{\partial t} p_{\ell'}^n(x) dx + [\mathbf{f}^R \cdot p_{\ell'}^n(x)]_{x_{n-1}}^{x_n} - \int_{x_{n-1}}^{x_n} \mathbf{f} \frac{\partial p_{\ell'}^n}{\partial x} dx = 0.$$

At any interface between two elements, i.e., the end points  $x_{n-1}$  and  $x_n$ , the flux vector  $\mathbf{f}^R$  is not uniquely determined and a flux formula has to be supplied to complete the discretization process. Various kinds of flux formulas have been proposed and used in the literature. In this paper, we will consider two commonly used flux formulas. They are specified below and will be referred to as the *characteristics-based* flux formula and *Lax-Friedrich* flux formula respectively.

The characteristics-based flux formula is of the form

$$(2.6) \quad \mathbf{f}^R(\mathbf{u}_L, \mathbf{u}_R) = \frac{1}{2}[\mathbf{f}(\mathbf{u}_L) + \mathbf{f}(\mathbf{u}_R) - \theta|\mathbf{A}|(\mathbf{u}_R - \mathbf{u}_L)], \quad \theta \geq 0$$

where  $\mathbf{u}_L$  and  $\mathbf{u}_R$  are the values of  $\mathbf{u}$  at the interface calculated using expansion coefficients of the elements at the left and right of that interface respectively. (In DGM  $\mathbf{u}_L$  and  $\mathbf{u}_R$  are not required to be the same.) Here  $\theta$  is a scalar parameter. The value of  $\theta$  is usually unity in practice which makes (2.6) an *exact characteristics splitting* (the exact Roe solver). On the other hand, (2.6) will result in a symmetric averaged scheme when  $\theta = 0$ . Here, we will keep  $\theta$  as a parameter so that our analysis can be useful for a wide range of cases. For convenience, (2.6) will be written as

$$(2.7) \quad \mathbf{f}^R(\mathbf{u}_L, \mathbf{u}_R) = \mathbf{A}_L \mathbf{u}_L + \mathbf{A}_R \mathbf{u}_R$$

where

$$(2.8) \quad \mathbf{A}_L = \frac{1}{2}[\mathbf{A} + \theta|\mathbf{A}|], \quad \mathbf{A}_R = \frac{1}{2}[\mathbf{A} - \theta|\mathbf{A}|].$$

The Lax-Friedrich flux formula is of the form

$$(2.9) \quad \mathbf{f}^R(\mathbf{u}_L, \mathbf{u}_R) = \frac{1}{2}[\mathbf{f}(\mathbf{u}_L) + \mathbf{f}(\mathbf{u}_R) - \theta|a|_{max}(\mathbf{u}_R - \mathbf{u}_L)] \quad \theta \geq 0$$

where  $|a|_{max}$  is the maximum (absolute value) of the eigenvalues of  $\mathbf{A}$ . This can again be written in the form of (2.7) with

$$(2.10) \quad \mathbf{A}_L = \frac{1}{2}[\mathbf{A} + \theta|a|_{max}\mathbf{I}], \quad \mathbf{A}_R = \frac{1}{2}[\mathbf{A} - \theta|a|_{max}\mathbf{I}].$$

Using expression (2.7) for both cases, the semi-discrete equation (2.5) can now be written as

$$(2.11) \quad \int_{x_{n-1}}^{x_n} \frac{\partial \mathbf{u}_h^n}{\partial t} p_{\ell'}^n(x) dx + [\mathbf{A}_L \mathbf{u}_h^n(x_n, t) + \mathbf{A}_R \mathbf{u}_h^{n+1}(x_n, t)] p_{\ell'}^n(x_n) \\ - [\mathbf{A}_L \mathbf{u}_h^{n-1}(x_{n-1}, t) + \mathbf{A}_R \mathbf{u}_h^n(x_{n-1}, t)] p_{\ell'}^n(x_{n-1}) - \int_{x_{n-1}}^{x_n} \mathbf{A} \mathbf{u}_h^n \frac{\partial p_{\ell'}^n}{\partial x} dx = 0$$

for  $\ell' = 0, 1, \dots, p$ .

Together with (2.3), equation (2.11) yields a system of time evolution equations for the expansion coefficients for each element. This system is usually solved by some time integration scheme such as the Runge-Kutta schemes ([2], [7]).

### 3. Spatially propagating waves in uniform grids.

**3.1. Use of local variables.** Introducing a local coordinate  $\xi$  for each element, we let

$$(3.1) \quad \xi = \frac{2}{\Delta x_n}(x - \bar{x}_n) \quad \text{where} \quad \Delta x_n = x_n - x_{n-1} \quad \text{and} \quad \bar{x}_n = \frac{x_{n-1} + x_n}{2}.$$

In addition, the basis functions will be taken to be the same for all the elements when expressed in the local coordinate  $\xi$ , i.e., we assume

$$p_\ell^n(x) = P_\ell(\xi),$$

where  $\{P_\ell(\xi), \ell = 0, 1, \dots, p\}$  is a chosen set of basis functions, such as the Legendre polynomials, or the set of  $\{1, \xi, \xi^2, \dots, \xi^p\}$ . The results of our analysis are independent of the specific choice on the basis functions.

We look for wave-like solutions supported by (2.11). By assuming a periodicity in time with a frequency  $\omega$ , we let

$$(3.2) \quad \mathbf{u}_h^n(\xi, t) = e^{-i\omega t} \hat{\mathbf{u}}_h^n(\xi) \quad \text{where} \quad \hat{\mathbf{u}}_h^n(\xi) = \sum_{\ell=0}^p \hat{\mathbf{c}}_\ell^n P_\ell(\xi), \quad \text{and} \quad i = \sqrt{-1}$$

The expansion coefficients  $\hat{\mathbf{c}}_\ell^n$  are now independent of  $t$ . Substituting the above into (2.11), we get

$$(3.3) \quad \begin{aligned} & -\frac{i\omega\Delta x_n}{2} \int_{-1}^1 \hat{\mathbf{u}}_h^n(\xi) \cdot P_{\ell'}(\xi) d\xi + [\mathbf{A}_L \hat{\mathbf{u}}_h^n(1) + \mathbf{A}_R \hat{\mathbf{u}}_h^{n+1}(-1)] P_{\ell'}(1) \\ & - [\mathbf{A}_L \hat{\mathbf{u}}_h^{n-1}(1) + \mathbf{A}_R \hat{\mathbf{u}}_h^n(-1)] P_{\ell'}(-1) - \int_{-1}^1 \mathbf{A} \hat{\mathbf{u}}_h^n(\xi) \frac{\partial P_{\ell'}}{\partial \xi} d\xi = 0 \end{aligned}$$

for  $\ell' = 0, 1, \dots, p$ .

**3.2. Uniform grid and the eigenvalue problem.** We now consider the case where elements are uniform in length, i.e.,  $\Delta x_n \equiv h$ . After substituting (2.3) into (3.3), we look for solutions with expansion coefficients of the form

$$(3.4) \quad \hat{\mathbf{c}}_\ell^n = \lambda^n \tilde{\mathbf{c}}_\ell$$

where  $\lambda$  is an undetermined complex number and  $\tilde{\mathbf{c}}_\ell$  is a vector independent of the element index  $n$ . It is easy to see that, if we express  $\lambda$  as

$$(3.5) \quad \lambda = e^{ik_h h},$$

then,  $k_h$  can be interpreted as the wave number of the numerical solution. Here  $k_h$  will be referred to as the *numerical wave number*.

For convenience of discussion, we define a column vector that contains all the expansion coefficients

$$(3.6) \quad \tilde{\mathbf{x}} = \begin{bmatrix} \tilde{\mathbf{c}}_0 \\ \tilde{\mathbf{c}}_1 \\ \dots \\ \tilde{\mathbf{c}}_p \end{bmatrix}$$



and matrices

$$(3.7) \quad \mathbf{Q} = \{q_{\ell'\ell}\}, \quad \text{where} \quad q_{\ell'\ell} = \int_{-1}^1 P_{\ell}(\xi) P_{\ell'}(\xi) d\xi,$$

$$(3.8) \quad \mathbf{Q}' = \{q'_{\ell'\ell}\}, \quad \text{where} \quad q'_{\ell'\ell} = \int_{-1}^1 P_{\ell}(\xi) \frac{\partial P_{\ell'}}{\partial \xi} d\xi,$$

$$(3.9) \quad \mathbf{B}_{(a,b)} = \{b_{\ell'\ell}\}, \quad \text{where} \quad b_{\ell'\ell} = P_{\ell'}(a) P_{\ell}(b)$$

where  $\ell', \ell = 0, 1, \dots, p$ . Then, equation (3.3) can be re-written compactly as

$$(3.10) \quad -\frac{i\omega h}{2}(\mathbf{Q} \otimes \mathbf{I})\tilde{\mathbf{x}} + (\mathbf{B}_{(1,1)} \otimes \mathbf{A}_L)\tilde{\mathbf{x}} - (\mathbf{B}_{(-1,-1)} \otimes \mathbf{A}_R)\tilde{\mathbf{x}} - (\mathbf{Q}' \otimes \mathbf{A})\tilde{\mathbf{x}} \\ + \lambda(\mathbf{B}_{(1,-1)} \otimes \mathbf{A}_R)\tilde{\mathbf{x}} - \frac{1}{\lambda}(\mathbf{B}_{(-1,1)} \otimes \mathbf{A}_L)\tilde{\mathbf{x}} = 0$$

where  $\otimes$  denotes the Kronecker product (The definition and relevant properties of  $\otimes$  can be found in appendix A1). For a given frequency  $\omega$ , equation (3.10) forms an eigenvalue problem with  $\lambda$  being the eigenvalue and  $\tilde{\mathbf{x}}$  the eigenvector.

Next, we show that (3.10) can be equivalently separated into  $N$  independent eigenvalue problems, where each sub-problem corresponds to one of the physical wave modes of the PDE. Here each pair of the eigenvalue and eigenvector  $\{a_j, \mathbf{e}_j\}$  of the PDE will be referred to as a *wave mode* of the PDE (2.1) and  $a_j$  is the wave speed of that mode. Since we are interested in spatially propagating waves, we assume  $a_j \neq 0$ .

We first express  $\tilde{\mathbf{x}}$  given in (3.6) in terms of eigenvectors of the PDE as follows,

$$(3.11) \quad \tilde{\mathbf{x}} = \begin{bmatrix} \tilde{\mathbf{c}}_0 \\ \tilde{\mathbf{c}}_1 \\ \dots \\ \tilde{\mathbf{c}}_p \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^N y_{0j} \mathbf{e}_j \\ \sum_{j=1}^N y_{1j} \mathbf{e}_j \\ \dots \\ \sum_{j=1}^N y_{pj} \mathbf{e}_j \end{bmatrix} = \sum_{j=1}^N \begin{bmatrix} y_{0j} \\ y_{1j} \\ \dots \\ y_{pj} \end{bmatrix} \otimes \mathbf{e}_j \equiv \sum_{j=1}^N \mathbf{y}_j \otimes \mathbf{e}_j$$

where  $\mathbf{y}_j$  is a column vector of dimension  $p+1$ . By substituting (3.11) into (3.10) and using a property of the Kronecker product (equation (7.2) of Appendix A1), we get

$$(3.12) \quad \sum_{j=1}^N \left( -\frac{i\omega h}{2}(\mathbf{Q}\mathbf{y}_j) \otimes \mathbf{e}_j + (\mathbf{B}_{(1,1)}\mathbf{y}_j) \otimes \mathbf{A}_L\mathbf{e}_j - (\mathbf{B}_{(-1,-1)}\mathbf{y}_j) \otimes \mathbf{A}_R\mathbf{e}_j - (\mathbf{Q}'\mathbf{y}_j) \otimes \mathbf{A}\mathbf{e}_j \right. \\ \left. + \lambda(\mathbf{B}_{(1,-1)}\mathbf{y}_j) \otimes \mathbf{A}_R\mathbf{e}_j - \frac{1}{\lambda}(\mathbf{B}_{(-1,1)}\mathbf{y}_j) \otimes \mathbf{A}_L\mathbf{e}_j \right) = 0.$$

Furthermore, for  $\mathbf{A}_L$  and  $\mathbf{A}_R$  given in (2.8) and (2.10), we have

$$(3.13) \quad \mathbf{A}_L\mathbf{e}_j = \frac{1 + \theta\beta_j}{2}a_j\mathbf{e}_j, \quad \mathbf{A}_R\mathbf{e}_j = \frac{1 - \theta\beta_j}{2}a_j\mathbf{e}_j,$$

where

$$(3.14) \quad \beta_j = \frac{|a_j|}{a_j}$$

for the characteristics-based flux defined in (2.8) and

$$(3.15) \quad \beta_j = \frac{|a|_{max}}{a_j}$$

for the Lax-Friedrich flux defined in (2.10). Thus the two types of flux formulas can be treated in a unified way by using (3.13). For convenience of discussion, we define

$$(3.16) \quad \gamma_j = \theta \beta_j.$$

and  $\gamma_j$  will be referred to as the *upwind factor* of the scheme for the  $j$ th wave mode of the PDE. We note that, for both cases given in (3.14) and (3.15),  $|\gamma_j| = 1$  leads to the exact characteristics splitting. For  $|\gamma_j| > 1$ , the eigenvalues of  $\mathbf{A}_L$  and  $\mathbf{A}_R$  are all positive and negative, respectively. Note also that for the Lax-Friedrich flux formula applied to a system of equations in which the  $|a_j|$  varies widely,  $|\gamma_j|$  will be large for the slowest wave modes; however, the wavenumber  $\omega h/a_j$  will also be proportionally large for any given  $\omega$ .

Equation (3.12) can now be expressed as

$$(3.17) \quad \sum_{j=1}^N \left( -\frac{i\omega h}{2} \mathbf{Q} \mathbf{y}_j + \frac{1+\gamma_j}{2} a_j \mathbf{B}_{(1,1)} \mathbf{y}_j - \frac{1-\gamma_j}{2} a_j \mathbf{B}_{(-1,-1)} \mathbf{y}_j - a_j \mathbf{Q}' \mathbf{y}_j + \lambda \frac{1-\gamma_j}{2} a_j \mathbf{B}_{(1,-1)} \mathbf{y}_j - \frac{1}{\lambda} \frac{1+\gamma_j}{2} a_j \mathbf{B}_{(-1,1)} \mathbf{y}_j \right) \otimes \mathbf{e}_j = 0.$$

Due to linear independency of  $\mathbf{e}_j$ 's, it is easy to see that (3.17) yields  $N$  independent sub-eigenvalue problems,

$$(3.18) \quad \left[ -\frac{i\omega h}{a_j} \mathbf{Q} + (1+\gamma_j) \mathbf{B}_{(1,1)} - (1-\gamma_j) \mathbf{B}_{(-1,-1)} - 2\mathbf{Q}' + \lambda(1-\gamma_j) \mathbf{B}_{(1,-1)} - \frac{1}{\lambda}(1+\gamma_j) \mathbf{B}_{(-1,1)} \right] \mathbf{y}_j = 0$$

for  $j = 1, 2, \dots, N$ , in which  $\mathbf{y}_j$  is the eigenvector and  $\lambda$  is the eigenvalue.

We observe that, by solving the eigenvalue problem posed in (3.18), we will obtain  $\lambda$  as a function of the non-dimensional frequency  $\frac{\omega h}{a_j}$  (or wavenumber) and the upwind factor  $\gamma_j$ , i.e.,

$$(3.19) \quad \lambda = F\left(\frac{\omega h}{a_j}, \gamma_j\right).$$

Since  $\lambda$  is directly related to the numerical wave number  $k_h$  by (3.5), equation (3.19) is the numerical dispersion relation of the scheme. It is an intrinsic property of the discretization.

In addition, the non-trivial solution of (3.18) forms the eigenfunction of the numerical mode. Specifically, let the eigenvectors of (3.18) be denoted by  $\mathbf{y} = \{\mathbf{v}_\ell\}$ , then the eigenfunctions will be of the form

$$(3.20) \quad \mathbf{u}_h^n(\xi, t) = e^{-i\omega t} \hat{u}_h^n(\xi) \mathbf{e}_j,$$

where

$$(3.21) \quad \hat{u}_h^n(\xi) = e^{in k_h h} f(\xi; \frac{\omega h}{a_j}, \gamma_j), \quad \text{and} \quad f(\xi; \frac{\omega h}{a_j}, \gamma_j) = \sum_{\ell=0}^p v_\ell P_\ell(\xi)$$

For convenience, the eigenfunctions will be normalized such that

$$(3.22) \quad f(1; \frac{\omega h}{a_j}, \gamma_j) = 1.$$

#### 4. Numerical dispersion relation.

**4.1. Determinant of equation (3.18).** Wave propagation properties of the numerical scheme are encoded in the numerical dispersion relation (3.19). For convenience of discussion, the subscript  $j$  will be dropped in this section. We define

$$K = \frac{\omega h}{a} \quad \text{and} \quad K_h = k_h h,$$

where  $K$  is the non-dimensional exact wave number of the PDE and  $K_h$  is the non-dimensional numerical wave number as given in (3.5).  $K_h$  is related to  $\lambda$  of (3.18) by

$$(4.1) \quad \lambda = e^{iK_h}.$$

By letting the determinant of the coefficient matrix for  $\mathbf{y}_j$  in (3.18) be zero, we get an algebraic equation for  $\lambda$  for any given value of  $K$ . We have computed the determinant of (3.18) symbolically using the computer algebra system MAPLE [16]. It is found that the determinant, after some normalization, can be written in the following form:

$$(4.2) \quad (1 - \gamma)[G(iK)\lambda - H(iK)] + (-1)^{p+1}(1 + \gamma)[G(-iK)\frac{1}{\lambda} - H(-iK)] = 0$$

where  $G(x)$  and  $H(x)$  are polynomials of degree  $p$  and  $p + 1$ , respectively, with real coefficients. The exact expressions for  $G$  and  $H$  are given in appendix A2.

Before presenting the numerical and analytical results of (4.2), we make some general remarks:

1. We observe that, due to the fact that the rank of  $\mathbf{B}$  matrices in (3.18) is unity (see equation (3.9)), equation (4.2) is quadratic in  $\lambda$ . Consequently, there will be at most two distinct solutions for  $\lambda$  in (4.2). Furthermore, when  $\gamma = \pm 1$  (exact characteristics splitting of the flux), equation (4.2) becomes linear in  $\lambda$  and there will be only one solution for  $\lambda$ .

2. In previous works in the literature, the spatial wave number  $K_h$ , thus  $\lambda$  in (4.2), is specified and  $K$  is to be solved from (4.2). This, of course, will result in solving a polynomial of degree  $p + 1$  which is difficult to do analytically for  $p > 3$ .

3. If  $\lambda = \lambda_0$  is a solution to (4.2) for  $\gamma = \gamma_0$ , then  $\lambda = \frac{1}{\lambda_0}$  is a solution for  $\gamma = -\gamma_0$ . (This can be shown simply by taking a complex conjugate of (4.2).) Thus, it is sufficient to consider theoretically only the cases with  $\gamma > 0$  in (4.2), i.e., right-going waves, for dissipation and dispersion errors.

**4.2. Numerical results of the eigensolutions.** We first present numerical results of (4.2). Its analytical properties will be presented in section 4.3. As we have seen in previous discussions, when  $|\gamma| \neq 1$ , there are two roots for  $\lambda$ . Each root represents a numerical wave mode whose wave number is found by (4.1) and whose mode shape (eigenfunctions) found by (3.21). As we will see, one of the numerical modes can faithfully represent the physical wave and the other mode is spurious or non-physical. These two modes behave very differently and it is easy to distinguish the physical mode from the non-physical mode. When  $|\gamma| = 1$  (exact characteristics flux), of course, there will be only one root and the spurious mode will not be present.

We will use a case with  $\gamma = 0.5$  as an example to demonstrate numerical results. For a given value of exact wave number  $K$ , we solve equation (4.2) and obtain two values of  $\lambda$  which are then converted into numerical wave numbers  $K_h$  according to (4.1). For the physical mode, the numerical wave number  $K_h$  as a function of  $K$  is plotted in Figure 4.1(a) (real part) and 4.1(b) (imaginary part), for cases  $p = 1, 2, 3, 4$ . The diagonal line in Figure 4.1(a) is the exact relation, i.e., when the numerical wave number is equal to the actual physical wave number. It is seen that, for a given value of  $p$ , the real part of the numerical wave number,  $Re(K_h)$ , follows the exact line closely for a range of  $K$  values. This range will be termed *resolved wave number space*. Clearly, the higher the order of the basis functions, the larger the resolved space. Figure 4.1(b) shows the imaginary part of the numerical wave number,  $Im(K_h)$ . Since the wave is right-traveling for the present case ( $\gamma > 0$ ), the positive imaginary part represents numerical damping as the wave propagates in space. We note that the damping is not significant for wave numbers within the resolved wave number space in each scheme. The exact boundary of resolved range is, of course, somewhat arbitrary and depends on the accuracy criteria imposed. This issue will be closely examined in section 5.2. In general, the dissipation error places a higher requirement on the resolution of the scheme than the dispersion error in DGM.

For the spurious mode, the relation of  $K_h$  v.s.  $K$  is plotted in Figure 4.2. For the real part of  $K_h$  shown in Figure 4.2(a), the curve starts at 0 for  $p$  odd and starts at  $\pi$  for  $p$  even. The group velocity of these waves (slope of  $Re(K_h)$  v.s.  $K$ ) is negative, indicating that the spurious waves are left-traveling, in the opposite direction of the actual physical wave. The imaginary part of  $K_h$  is also negative, indicating again that the wave is left-traveling and damped. The damping rates for the spurious modes in Figure 4.2(b) are quite large for the cases shown. This means that the spurious mode is expected to be damped very rapidly in computation.

The corresponding eigenfunctions of the physical and spurious modes are plotted in Figures 4.3 and 4.4 respectively. The eigenfunctions are constructed according to (3.21) using eigenvectors from (3.18) as the expansion coefficients. Plotted are eigenfunctions over a span of 30 elements, with the first element being  $[-1, 1]$  as indicated by dark lines in the plots. As shown in Figure 4.3, the physical mode travels to the right and the amount of damping is quite visible for  $p = 1$  and 2 with the chosen value of  $K = 2$ . The damping error reduces significantly as order increases.

In Figure 4.4, we see that the spurious non-physical mode is damped very rapidly for all cases shown, which is consistent with our observation in Figure 4.2.

As will be shown later (equation (40)), the damping factor of the spurious mode is related to the value of  $\gamma$  as  $\left| \frac{1-\gamma}{1+\gamma} \right|$ , plotted in Figure 4.5. Thus the spurious wave modes become highly damped when  $\gamma$  is close to unity and much less damped when  $\gamma$  is close to zero or much greater than unity. In practice, small  $\gamma$  is avoided by choosing  $\theta \sim 1$ ; large  $\gamma$  occurs for slow wave modes where  $|a|_{max}/a_j$  is small.

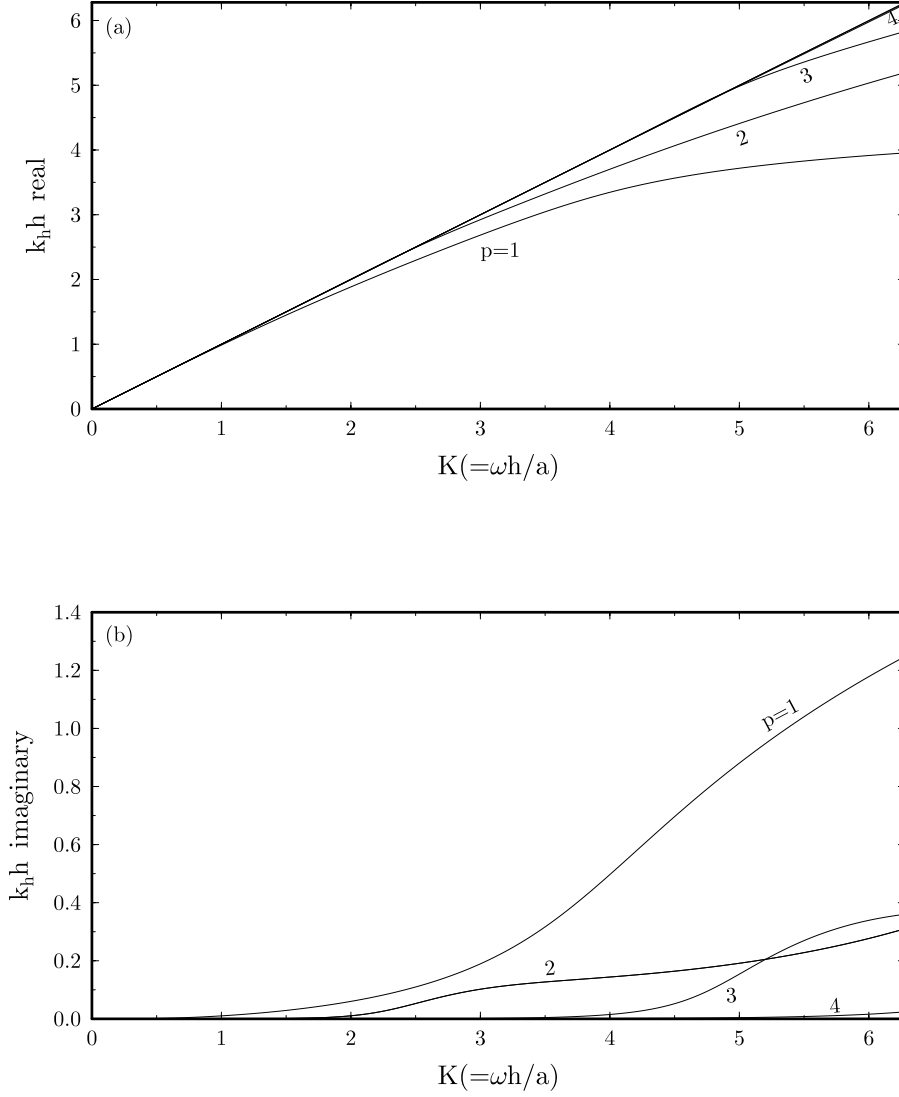


FIG. 4.1. The physical mode, numerical wave number  $k_h h$  v.s. actual wave number (normalized frequency)  $K = \frac{\omega h}{a}$ .  $\gamma = 0.5$ .

**4.3. Super-accuracy of the numerical wave number.** The numerical wave number  $K_h$  of the physical mode should be a close approximation of the actual wave number  $K$ , especially in the long wavelength limit, i.e., when  $K$  is small. Here, we give an estimation on the order of convergence in wave number space and show that  $K_h$  is accurate to the actual wave number  $K$  to order  $2p + 2$ , which is twice the order of accuracy of the basis functions.

Assuming  $|\gamma| \neq 1$ , we can re-write (4.2) as a quadratic equation for  $\lambda$  as follows,

$$(4.3) \quad \lambda^2 - \left[ \frac{H(iK)}{G(iK)} + (-1)^{p+1} \frac{1 + \gamma}{1 - \gamma} \frac{H(-iK)}{G(iK)} \right] \lambda + (-1)^{p+1} \frac{1 + \gamma}{1 - \gamma} \frac{G(-iK)}{G(iK)} = 0.$$

By examining the computed  $H(x)$  and  $G(x)$  functions (appendix A2), we found that the ratio  $H(x)/G(x)$

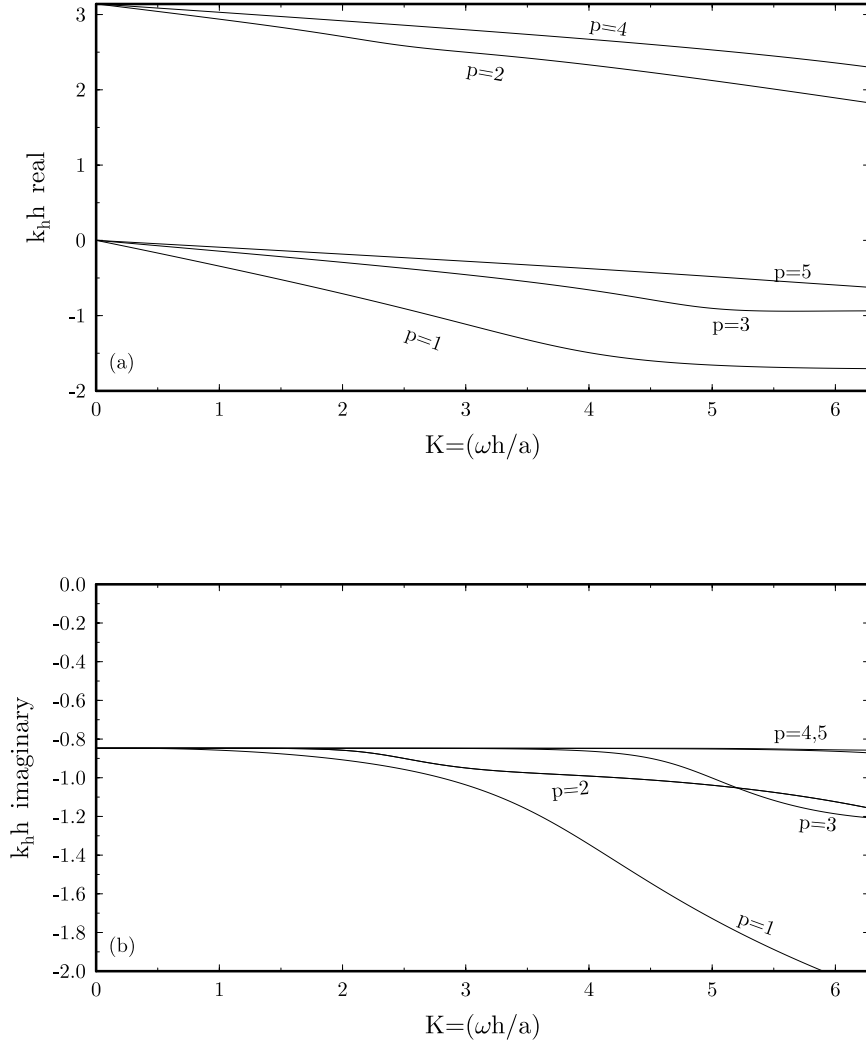


FIG. 4.2. The spurious mode, numerical wave number  $k_h h$  v.s. actual wave number (normalized frequency)  $K = \frac{\omega h}{a}$ .  $\gamma = 0.5$ .

is always *exactly* the Pade approximation of  $e^x$  to order  $2p + 2$ . (This has been calculated and verified symbolically for  $p$  up to 16 and is conjectured to be true for all  $p$ .) That is, we have

$$(4.4) \quad \frac{H(x)}{G(x)} = e^x + O(x^{2p+2}).$$

Consequently, we can show that, for  $K$  small, the two roots of (4.3) are

$$(4.5) \quad \lambda^{(p)} = e^{iK} + C_1 (iK)^{2p+2} + C_2 (iK)^{2p+3} + \dots \quad (\text{physical mode})$$

and

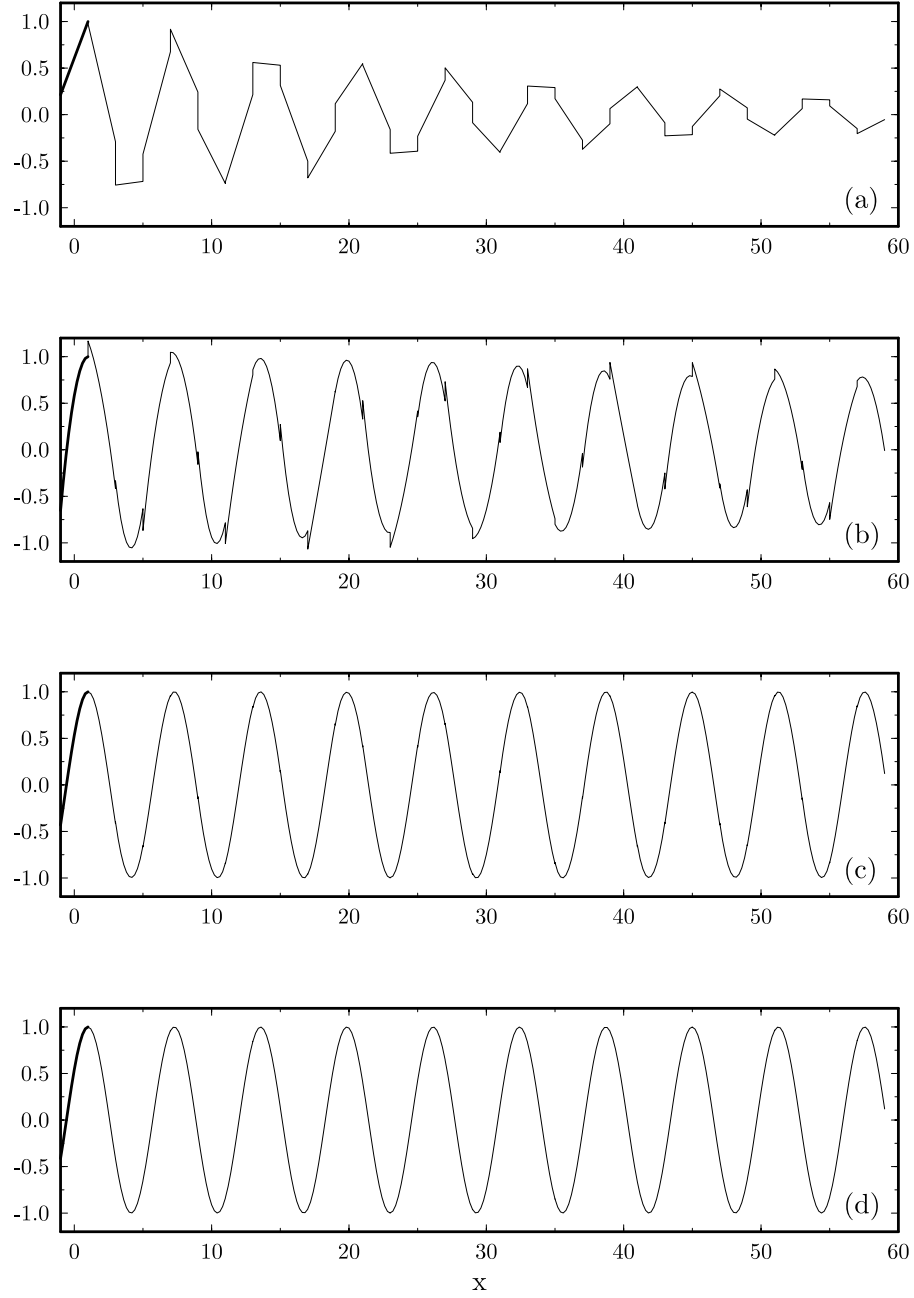


FIG. 4.3. *Eigen-functions of the physical mode.*  $K = 2$ .  $\gamma = 0.5$ . (a)  $p = 1$ , (b)  $p = 2$ , (c)  $p = 3$ , (d)  $p = 4$ .

$$(4.6) \lambda^{(s)} = (-1)^{p+1} \frac{1 + \gamma}{1 - \gamma} \frac{G(-iK)}{G(iK)} e^{-iK} + D_1 (iK)^{2p+2} + D_2 (iK)^{2p+3} + \dots \quad (\text{non-physical mode})$$

where  $C_1$ ,  $C_2$  and  $D_1$ ,  $D_2$  are real coefficients, and dots represent higher order terms in  $(iK)$ . A detailed derivation is given in the appendix A3. Here the superscripts (p) and (s) indicate the physical and spurious modes respectively.

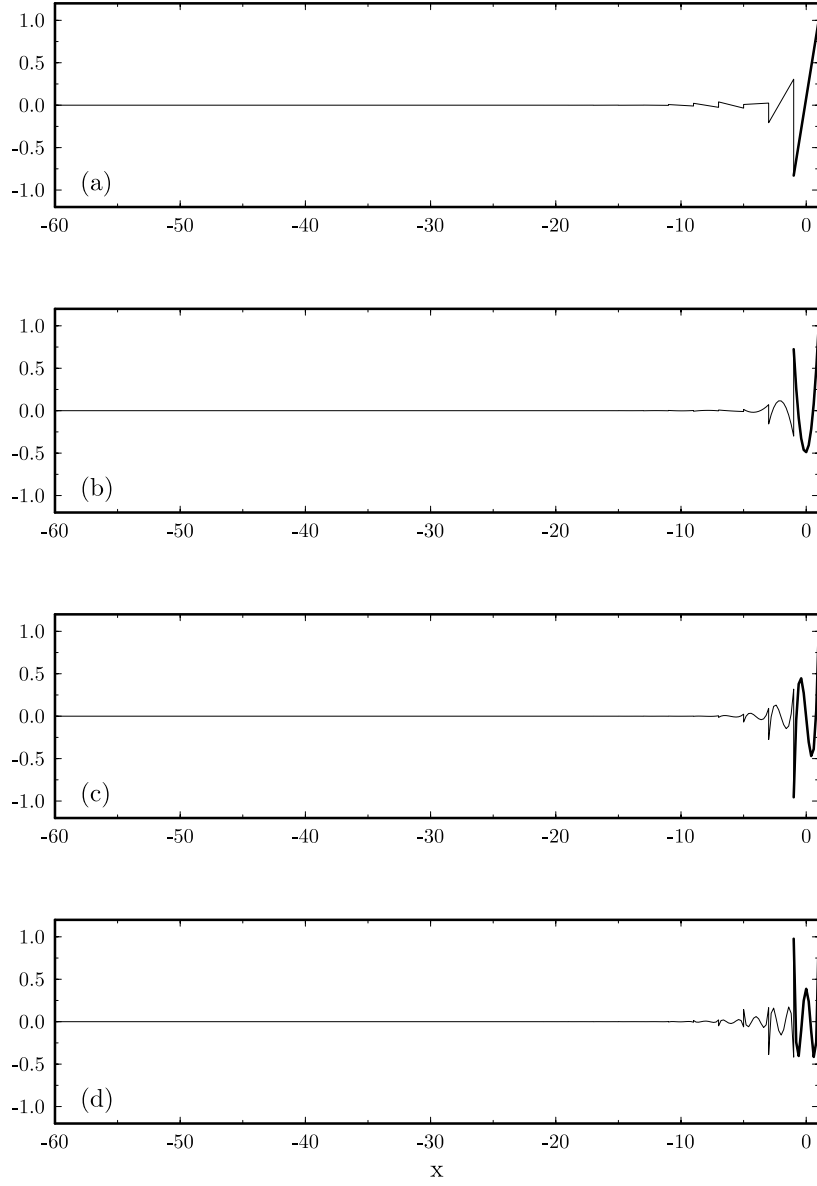


FIG. 4.4. *Eigen-functions of the spurious mode.*  $K = 2$ .  $\gamma = 0.5$ . (a)  $p = 1$ , (b)  $p = 2$ , (c)  $p = 3$ , (d)  $p = 4$ .

Thus, for the numerical wave number of the physical mode  $K_h^{(p)}$ , equation (4.5) gives

$$e^{iK_h^{(p)}} = e^{iK} + C_1(iK)^{2p+2} + C_2(iK)^{2p+3} + \dots$$

Therefore, we get the following order estimate

$$K_h^{(p)} = -i \ln[e^{iK} + C_1(iK)^{2p+2} + C_2(iK)^{2p+3} + \dots] = K - iC_1(iK)^{2p+2} - iC_2'(iK)^{2p+3} + \dots$$



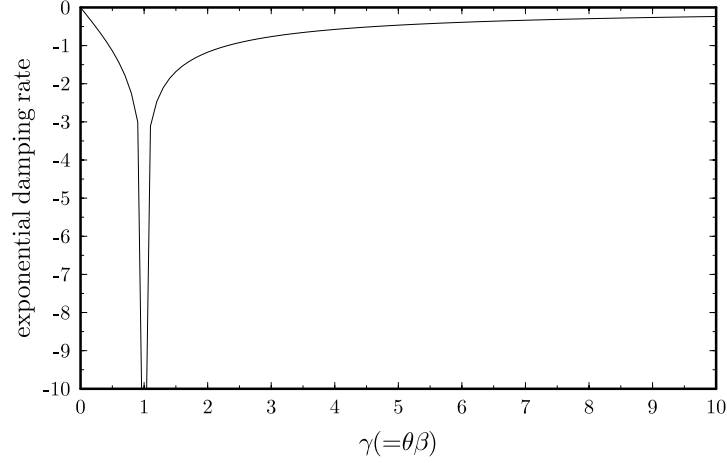


FIG. 4.5. *Effects of  $\gamma$  on the damping rate of the spurious mode.*

$$(4.7) \quad = K + (-1)^p i C_1 K^{2p+2} + (-1)^{p+1} C'_2 K^{2p+3} + \dots$$

where  $C'_2$  is also a real coefficient. Furthermore, by considering the real and imaginary parts of (4.7), we can get an estimation on the convergence rates of the dispersion and dissipation errors. Specifically, we have

$$(4.8) \quad \text{dispersion error :} \quad \text{Re}(K_h^{(p)}) - K = (-1)^{p+1} C'_2 K^{2p+3} + \dots,$$

$$(4.9) \quad \text{dissipation error :} \quad \text{Im}(K_h^{(p)}) = (-1)^p C_1 K^{2p+2} + \dots.$$

That is, for DGM, the dominant error is the dissipation error which reduces locally at order  $2p + 2$ . The dispersion error, on the other hand, reduces locally at order  $2p + 3$ . This is confirmed in Figure 4.6 where the numerical dispersion relations shown in Figure 4.1 are re-plotted in log-log scale.

Note that when  $|\gamma| = 1$ , the spurious mode is non-existent and it is straightforward to verify directly from (4.2) that (4.7)-(4.9) are still true for the physically accurate mode.

We also note that for polynomials  $H(x)$  and  $G(x)$  with given orders, (4.4) is the best possible order of approximation. This suggests that (4.7) is the best asymptotic numerical dispersion relation possible.

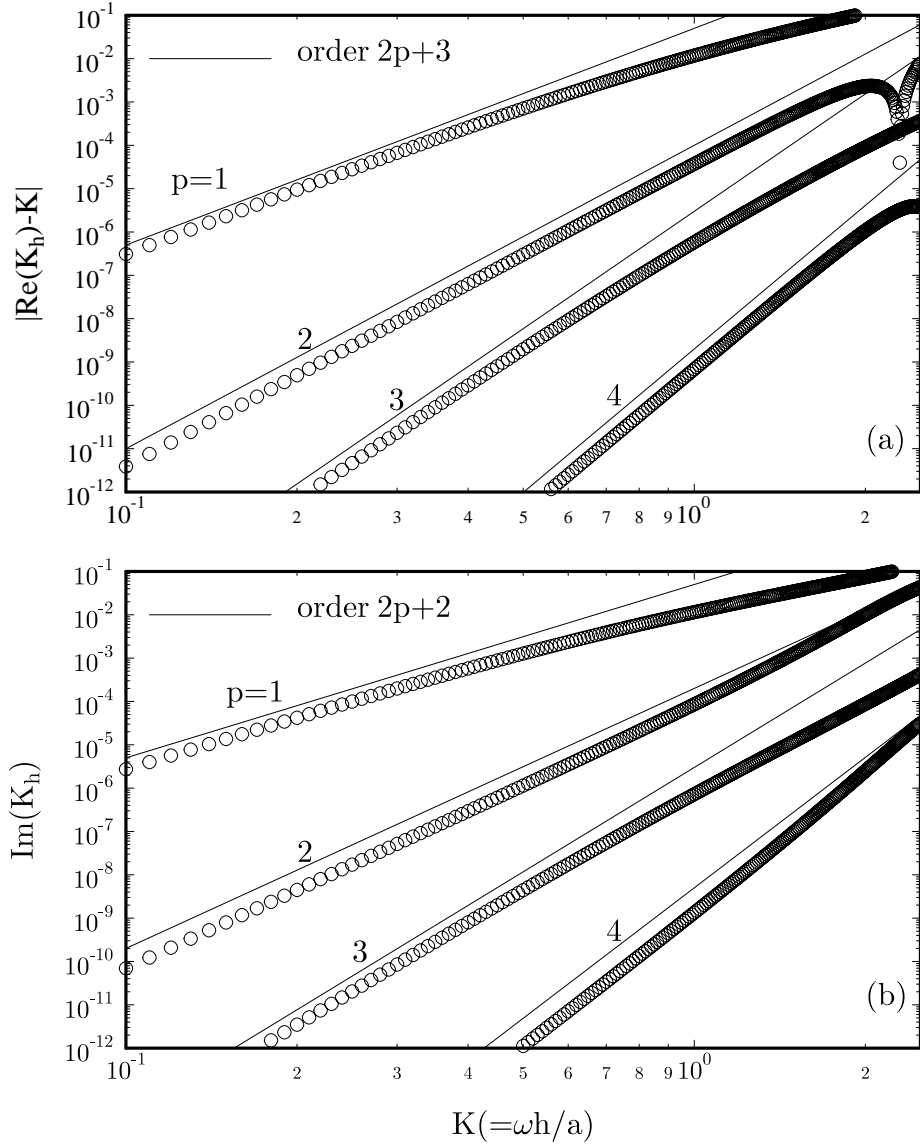


FIG. 4.6. Local order of convergence for the dispersion and dissipation errors of the physical mode. Circle: numerical wave number computed by equation (4.2); solid line: theoretical convergence rate. (a) dispersion error; (b) dissipation error.  $\gamma = 0.5$ .

## 5. Wave reflection at an interface of mesh discontinuity.

**5.1. Reflected and transmitted waves.** In this section, we consider a situation where the size of the element is abruptly changed from  $h_1$  to  $h_2$  across the interface between elements  $n = 0$  and  $n = 1$ , as shown in Figure 5.1. We will study the wave reflection and transmission at the interface. Specifically, we will introduce an incident physical wave, traveling from left to right, and look for the reflected and transmitted waves caused by the grid discontinuity.

Using the eigenfunction expression (3.20), we can express the incident wave as

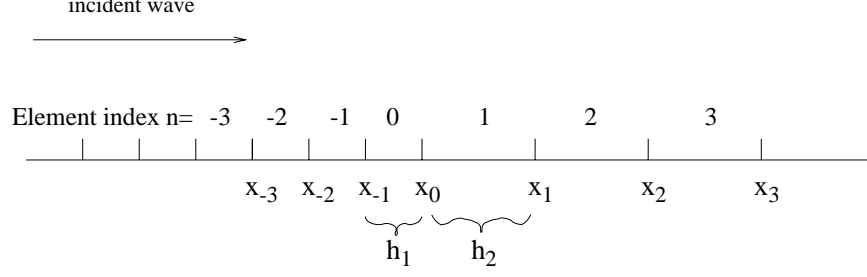


FIG. 5.1. A schematic of grid change.

$$(5.1) \quad \mathbf{u}_{incident} = A_0 e^{-i\omega t} e^{inK_{h_1}^{(p)} f_{j_0}^{(p)}(\xi; \frac{\omega h_1}{a_{j_0}}, \gamma_{j_0})} \mathbf{e}_{j_0}$$

where  $A_0$  is the wave amplitude. Here  $\mathbf{e}_{j_0}$  denotes the eigenvector of a right-going wave mode of the PDE (2.1), and  $K_{h_1, j_0}^{(p)}$  and  $f_{j_0}^{(p)}(\xi; \frac{\omega h_1}{a_{j_0}}, \gamma_{j_0})$  are the numerical wave number and eigenfunction for that wave mode found assuming a uniform mesh  $h_1$ . In other words, (5.1) satisfies the time harmonic semi-discrete equation (3.3) if  $\Delta x_n = h_1$  is held for all  $n$ . The superscript  $(p)$  in (5.1) denotes that the incident numerical mode is a physical wave mode. Likewise, a superscript  $(s)$  will be used to denote the spurious modes.

Due to the discontinuity in mesh size, there will be reflections at the interface. For convenience of discussion, let  $\hat{\mathbf{u}}_{left}^n$  and  $\hat{\mathbf{u}}_{right}^n$  denote the time independent solutions in the left and right half-domains on either side of the interface respectively. By making a use of (3.21), we get

$$(5.2) \quad \begin{aligned} \hat{\mathbf{u}}_{left}^n = & \underbrace{A_0 e^{inK_{h_1, j_0}^{(p)} f_{j_0}^{(p)}(\xi; \frac{\omega h_1}{a_{j_0}}, \gamma_{j_0})} \mathbf{e}_{j_0}}_{\text{incident}} \\ & + \underbrace{A_r e^{inK_{h_1, j_0}^{(s)} f_{j_0}^{(s)}(\xi; \frac{\omega h_1}{a_{j_0}}, \gamma_{j_0})} \mathbf{e}_{j_0} + \sum_{j \neq j_0} B_j e^{inK_{h_1, j}^{-} f_j^{-}(\xi; \frac{\omega h_1}{a_j}, \gamma_j)} \mathbf{e}_j}_{\text{reflected}} \end{aligned}$$

and

$$(5.3) \quad \hat{\mathbf{u}}_{right}^n = \underbrace{A_t e^{inK_{h_2, j_0}^{(p)} f_{j_0}^{(p)}(\xi; \frac{\omega h_2}{a_{j_0}}, \gamma_{j_0})} \mathbf{e}_{j_0} + \sum_{j \neq j_0} A_j e^{inK_{h_2, j}^{+} f_j^{+}(\xi; \frac{\omega h_2}{a_j}, \gamma_j)} \mathbf{e}_j}_{\text{transmitted}}$$

Here,  $A_r$  and  $A_t$  are amplitudes of the reflected and transmitted waves associated with the  $\mathbf{e}_{j_0}$  wave and  $B_j$  and  $A_j$  are those associated with the other waves of the PDE. The superscripts  $+$  and  $-$  in the terms inside the summations of (5.2) and (5.3) denote the direction of propagation (right-traveling and left-traveling respectively). It will be shown next, however, that all  $B_j$  and  $A_j$  are zero.

**5.2. Matching conditions at the interface.** To derive matching conditions at the interface, we first note that  $\hat{\mathbf{u}}_{left}^n$  and  $\hat{\mathbf{u}}_{right}^n$  satisfy equation (3.3) for a uniform element size  $h_1$  and  $h_2$  respectively. The coupling of the solutions can be found by applying (3.3) at the two adjacent elements near the interface of the grid discontinuity, namely, at elements  $n = 0$  and  $n = 1$ , Figure 5.1. Thus, from (3.3), we have

$n = 0$ :

$$(5.4) \quad -\frac{i\omega h_1}{2} \int_{-1}^1 \hat{\mathbf{u}}_{left}^0(\xi) \cdot P_{\ell'}(\xi) d\xi + [\mathbf{A}_L \hat{\mathbf{u}}_{left}^0(1) + \mathbf{A}_R \hat{\mathbf{u}}_{right}^1(-1)] P_{\ell'}(1) \\ - [\mathbf{A}_L \hat{\mathbf{u}}_{left}^{-1}(1) + \mathbf{A}_R \hat{\mathbf{u}}_{left}^0(-1)] P_{\ell'}(-1) - \int_{-1}^1 \mathbf{A} \hat{\mathbf{u}}_{left}^0(\xi) \frac{\partial P_{\ell'}}{\partial \xi} d\xi = 0$$

$n = 1$ :

$$(5.5) \quad -\frac{i\omega h_2}{2} \int_{-1}^1 \hat{\mathbf{u}}_{right}^1(\xi) \cdot P_{\ell'}(\xi) d\xi + [\mathbf{A}_L \hat{\mathbf{u}}_{right}^1(1) + \mathbf{A}_R \hat{\mathbf{u}}_{right}^2(-1)] P_{\ell'}(1) \\ - [\mathbf{A}_L \hat{\mathbf{u}}_{left}^0(1) + \mathbf{A}_R \hat{\mathbf{u}}_{right}^1(-1)] P_{\ell'}(-1) - \int_{-1}^1 \mathbf{A} \hat{\mathbf{u}}_{right}^1(\xi) \frac{\partial P_{\ell'}}{\partial \xi} d\xi = 0$$

These two conditions can be simplified when we recognize the fact that (5.4) and (5.5) will still be true when  $\hat{\mathbf{u}}_{right}^1(-1)$  in (5.4) is replaced by  $\hat{\mathbf{u}}_{left}^1(-1)$  and  $\hat{\mathbf{u}}_{left}^0(1)$  in (5.5) is replaced by  $\hat{\mathbf{u}}_{right}^0(1)$  due to the reason stated at the beginning of the section. Consequently, the matching conditions (5.4) and (5.5) are equivalent to the following two equations that are much more compact:

$$(5.6) \quad \mathbf{A}_R \hat{\mathbf{u}}_{left}^1(-1) = \mathbf{A}_R \hat{\mathbf{u}}_{right}^1(-1)$$

and

$$(5.7) \quad \mathbf{A}_L \hat{\mathbf{u}}_{left}^0(1) = \mathbf{A}_L \hat{\mathbf{u}}_{right}^0(1)$$

Now by substituting (5.2) and (5.3) into (5.6) and (5.7), and recalling (3.13), we easily get

$$(5.8) \quad A_0 e^{iK_{h_1,j_0}^{(p)}} f_{j_0}^{(p)}(-1; \frac{\omega h_1}{a_{j_0}}, \gamma_{j_0}) \frac{1 - \gamma_{j_0}}{2} a_{j_0} \mathbf{e}_{j_0} + A_r e^{iK_{h_1,j_0}^{(s)}} f_{j_0}^{(s)}(-1; \frac{\omega h_1}{a_{j_0}}, \gamma_{j_0}) \frac{1 - \gamma_{j_0}}{2} a_{j_0} \mathbf{e}_{j_0} \\ + \sum_{j \neq j_0} B_j e^{iK_{h_1,j}^-} f_j^-(-1; \frac{\omega h_1}{a_j}, \gamma_j) \frac{1 - \gamma_j}{2} a_j \mathbf{e}_j \\ = A_t e^{iK_{h_2,j_0}^{(p)}} f_{j_0}^{(p)}(-1; \frac{\omega h_2}{a_{j_0}}, \gamma_{j_0}) \frac{1 - \gamma_{j_0}}{2} a_{j_0} \mathbf{e}_{j_0} + \sum_{j \neq j_0} A_j e^{iK_{h_2,j}^+} f_j^+(-1; \frac{\omega h_2}{a_j}, \gamma_j) \frac{1 - \gamma_j}{2} a_j \mathbf{e}_j$$

and

$$A_0 \frac{1 + \gamma_{j_0}}{2} a_{j_0} \mathbf{e}_{j_0} + A_r \frac{1 + \gamma_{j_0}}{2} a_{j_0} \mathbf{e}_{j_0} + \sum_{j \neq j_0} B_j \frac{1 + \gamma_j}{2} a_j \mathbf{e}_j$$

$$(5.9) \quad = A_t \frac{1 + \gamma_{j_0}}{2} a_{j_0} \mathbf{e}_{j_0} + \sum_{j \neq j_0} A_j \frac{1 + \gamma_j}{2} a_j \mathbf{e}_j.$$

(In (5.9), we have used the fact that eigenfunctions are normalized such that  $f|_{\xi=1} = 1$ , as in (3.22).) Since  $\mathbf{e}_j$ 's are linearly independent, it follows that

$$(5.10) \quad B_j = A_j = 0, \quad j \neq j_0.$$

This means that no component of wave modes other than that of the incident mode  $\mathbf{e}_{j_0}$  will be present in the reflected and transmitted waves. This also suggests that the reflected wave can only be in the form of the spurious mode, the only opposite traveling numerical wave for the  $\mathbf{e}_{j_0}$  mode. An interesting consequence of this is that when the exact characteristics splitting flux formula is used, there will be no reflected wave because the opposite-traveling spurious wave is non-existent.

Further, by equating the coefficients of  $\mathbf{e}_{j_0}$  in (5.8) and (5.9), and assuming  $|\gamma_{j_0}| \neq 1$ , we get two coupled equations for  $A_r$  and  $A_t$ ,

$$A_0 e^{iK_{h_1, j_0}^{(p)} f_{j_0}^{(p)}(-1; \frac{\omega h_1}{a_{j_0}}, \gamma_{j_0})} + A_r e^{iK_{h_1, j_0}^{(s)} f_{j_0}^{(s)}(-1; \frac{\omega h_1}{a_{j_0}}, \gamma_{j_0})} = A_t e^{iK_{h_2, j_0}^{(p)} f_{j_0}^{(p)}(-1; \frac{\omega h_2}{a_{j_0}}, \gamma_{j_0})},$$

$$A_0 + A_r = A_t.$$

Solving the above, we get the following closed expressions for the reflection and transmission coefficients

$$(5.11) \quad \frac{A_r}{A_0} = \frac{e^{iK_{h_2, j_0}^{(p)} f_{j_0}^{(p)}(-1; \frac{\omega h_2}{a_{j_0}}, \gamma_{j_0})} - e^{iK_{h_1, j_0}^{(p)} f_{j_0}^{(p)}(-1; \frac{\omega h_1}{a_{j_0}}, \gamma_{j_0})}}{e^{iK_{h_1, j_0}^{(s)} f_{j_0}^{(s)}(-1; \frac{\omega h_1}{a_{j_0}}, \gamma_{j_0})} - e^{iK_{h_2, j_0}^{(p)} f_{j_0}^{(p)}(-1; \frac{\omega h_2}{a_{j_0}}, \gamma_{j_0})}}$$

$$(5.12) \quad \frac{A_t}{A_0} = \frac{e^{iK_{h_1, j_0}^{(s)} f_{j_0}^{(s)}(-1; \frac{\omega h_1}{a_{j_0}}, \gamma_{j_0})} - e^{iK_{h_1, j_0}^{(p)} f_{j_0}^{(p)}(-1; \frac{\omega h_1}{a_{j_0}}, \gamma_{j_0})}}{e^{iK_{h_1, j_0}^{(s)} f_{j_0}^{(s)}(-1; \frac{\omega h_1}{a_{j_0}}, \gamma_{j_0})} - e^{iK_{h_2, j_0}^{(p)} f_{j_0}^{(p)}(-1; \frac{\omega h_2}{a_{j_0}}, \gamma_{j_0})}}$$

Thus, numerical reflection and transmission coefficients are directly related to the change in dispersion properties of the scheme when grid change occurs.

To express the above in a more compact and, perhaps, more insightful form, we note the fact that numerical solutions in DGM have a small discontinuity (or gap) at the boundary of any two elements. This discontinuity, of course, becomes diminished with the increase of the resolution of the scheme. Specifically, if we let  $\Delta_h$  denote the discontinuity of the numerical solution at the interface of elements  $n = 0$  and  $n = 1$  had the grid size been uniformly  $h$ , then we have

$$\Delta_h = \hat{u}_h^1(-1) - \hat{u}_h^0(1) = e^{iK_h} f(-1; \frac{\omega h}{a}, \gamma) - 1$$

where  $u_h^n(\xi)$  is the eigenfunction specified in (3.21). Thus, in terms of  $\Delta_h$ , the expressions for the reflection and transmission coefficients given in (5.11) and (5.12) can now be written as

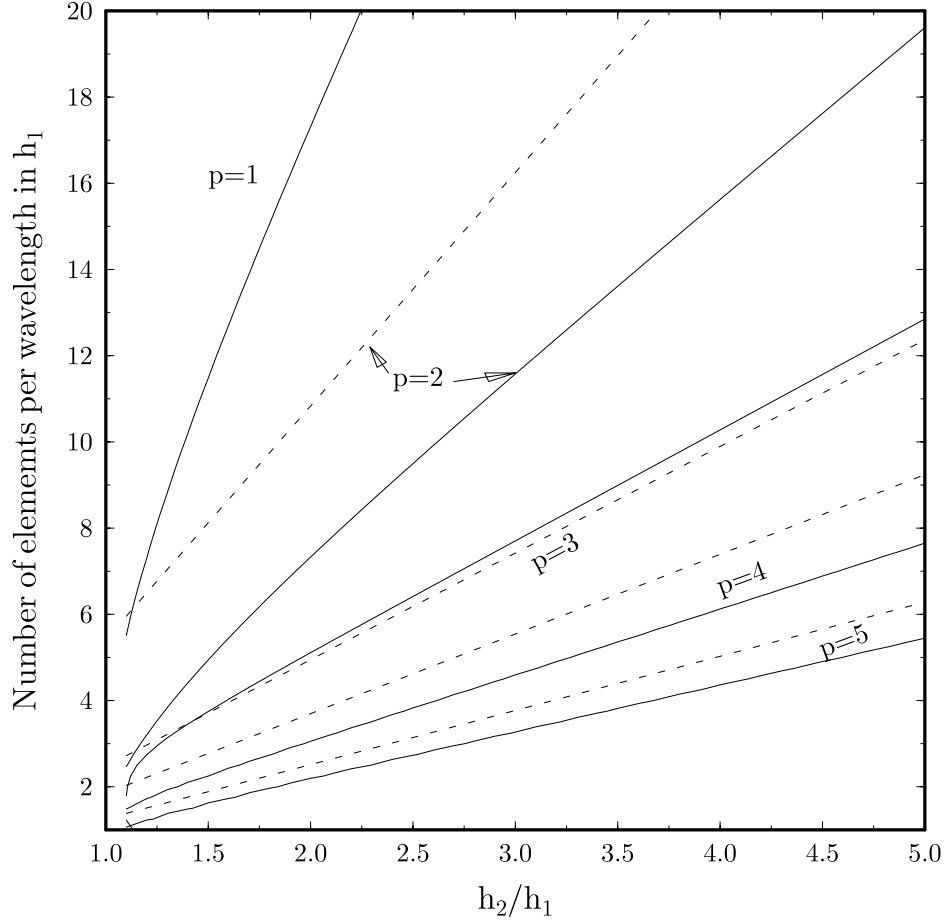


FIG. 5.2. Solid lines, boundaries of 2% numerical reflection. Reflected wave is less than 2% of the incident wave for parameters above the curves.  $\gamma = 0.5$ . Dashed lines, accuracy limits determined from the dispersion relation of a uniform grid  $h_2$ . The accuracy limit for  $p = 1$  (not shown) is far above and out of the picture.

$$(5.13) \quad \frac{A_r}{A_0} = \frac{\Delta_{h_2}^{(p)} - \Delta_{h_1}^{(p)}}{\Delta_{h_1}^{(s)} - \Delta_{h_2}^{(p)}}$$

$$(5.14) \quad \frac{A_t}{A_0} = \frac{\Delta_{h_1}^{(s)} - \Delta_{h_1}^{(p)}}{\Delta_{h_1}^{(s)} - \Delta_{h_2}^{(p)}}$$

in which the superscript denotes the mode type, the physical ( $p$ ) or spurious ( $s$ ) mode, and the subscript denotes the mesh spacing used for calculating the gap.

Equation (5.13) implies that numerical reflection will be small for waves that are well resolved under the grids on both sides of the interface, since the solution discontinuity decreases dramatically as the resolution of the scheme increases. This is further illustrated in Figure 5.2 where regions that satisfy the requirement on the resolution (number of elements per wavelength) so that the reflection is 2% or less are plotted for a given grid discontinuity of ratio  $h_2/h_1$ . A value of  $\gamma = 0.5$  is used in the calculations. The solid line is

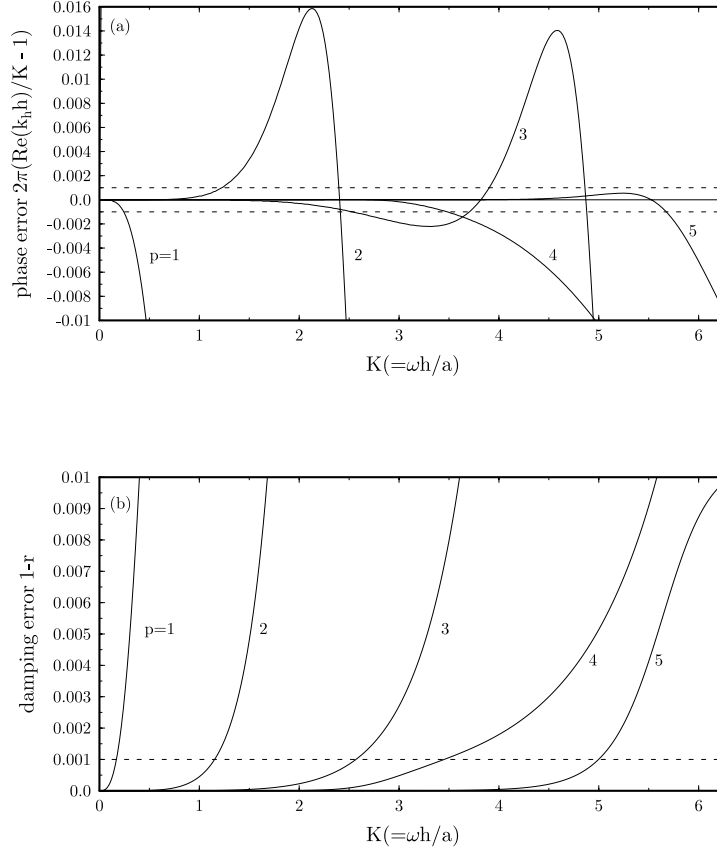


FIG. 5.3. Enlarged numerical phase (a) and damping (b) errors per wavelength of propagation for the physical mode.  $k_h$  is the numerical wave number and  $K$  is the exact wave number.  $r = e^{-2\pi \text{Im}(k_h h)/K}$  is the wave amplitude damping factor. The dotted lines indicate the accuracy limits used in plotting Figure 5.2.

the 2% reflection boundary for each given order of the scheme as indicated on the graph. As we can see, when the ratio  $h_2/h_1$  increases, the requirement on resolution also increases. It is interesting to compare this requirement with the resolution requirement placed by the accuracy criteria of the scheme had the grid been uniformly spaced. Since here we assume  $h_2 > h_1$ , the accuracy requirement will be calculated based on  $h_2$ . The accuracy boundaries are plotted in Figure 5.2 as dotted lines. The criteria used here consist of the dispersion error  $2\pi \frac{|\text{Re}(k_h h) - K|}{K} < 0.001$  and dissipation error  $1 - e^{-2\pi \text{Im}(k_h h)/K} < 0.001$ . This corresponds to requiring that the phase and damping errors be less than 10% after a wave has been propagated 100 wavelengths. Enlarged numerical dispersion relations are plotted in Figure 5.3 where the accuracy limits used are shown as dotted lines. Figure 5.3 indicates that the uniform grid accuracy constraint is similar to, and in many cases more stringent than, the accuracy constraint due to the abrupt change in mesh size. Although the uniform grid and discontinuous grid error criteria used here are somewhat arbitrary, we use Figure 5.2 to emphasize the notion that both types of errors follow parallel trends with respect to varying mesh sizes and the increase of the resolution of a scheme leads to the reduction of numerical reflection caused by a grid discontinuity.

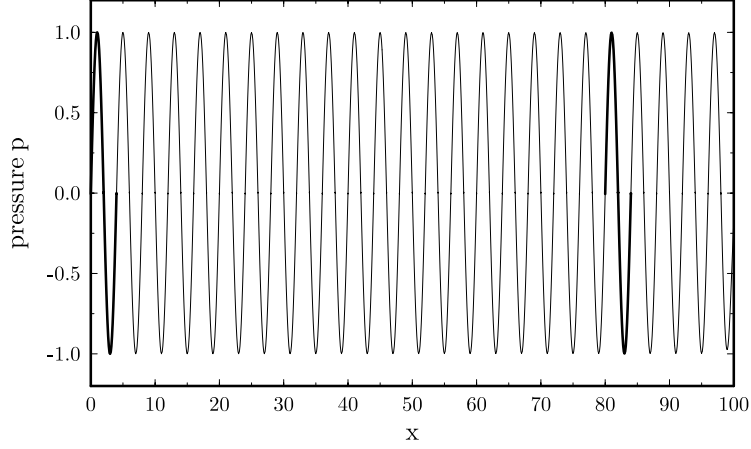


FIG. 6.1. Propagation of a periodic sine wave. The difference of the solutions at two periods shown in dark lines are computed in Table 6.1.

**6. Numerical examples.** In this section, we present numerical examples that illustrate and verify the wave propagation properties found in this paper.

**6.1. Super-accuracy of wave propagation.** We solve the linearized Euler equations with constant mean flow in 1-D:

$$(6.1) \quad \frac{\partial u}{\partial t} + M \frac{\partial u}{\partial x} + \frac{\partial p}{\partial x} = 0$$

$$(6.2) \quad \frac{\partial p}{\partial t} + M \frac{\partial p}{\partial x} + \frac{\partial u}{\partial x} = 0$$

where  $M$  is the mean flow Mach number,  $u$  is the velocity and  $p$  is the pressure. The Jacobian matrix has eigenvalues  $M-1$  and  $M+1$ , which represent the acoustic wave modes. We use Legendre polynomials as basis functions in our calculation. The semi-discrete equation is solved by an optimized 4th-order Runge-Kutta scheme (LDDRK56 in [10]).

To verify the accuracy of spatial propagation, we consider a computational domain of  $[0, 100]$  and introduce an incoming wave

$$(6.3) \quad \begin{bmatrix} u_{in} \\ p_{in} \end{bmatrix} = \sin[\omega_0(x-t)] \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

at the left boundary  $x = 0$ . The frequency is chosen to be  $\omega_0 = \pi/2$  with a wavelength  $\lambda_0 = 4$  in a mean flow  $M = 0$ . At the right boundary  $x = 100$ , we implement the characteristics boundary condition, i.e., the exact characteristics flux formula ( $\gamma = 1$ ) is used at the right boundary of the last element. After the initial transient has exited the right boundary, the computational domain is filled with the sine wave. We then



TABLE 6.1  
Solution of (6.1)-(6.2),  $M = 0$ , using uniform grids. Error is calculated by (6.4).

$p$	$h$	$\gamma = 1$		$\gamma = 0.5$	
		Error $E$	order	Error $E$	order
1	1	1.74054	-	1.79386	-
	0.5	1.09166	0.6730	1.46813	0.2890
	0.25	0.197915	2.4635	0.344971	2.0894
	0.125	0.0261657	2.9191	0.0506057	2.7691
2	1	0.27629	-	0.286715	-
	0.5	0.010116	4.7714	0.00634575	5.4976
	0.25	0.000323692	4.9658	0.000172082	5.2053
	0.125	$0.1016 \times 10^{-4}$	4.9923	$0.5165 \times 10^{-6}$	5.0574
3	1	0.00386958	-	0.00381781	-
	0.5	$0.3217 \times 10^{-4}$	6.9102	$0.4912 \times 10^{-4}$	6.2801
	0.25	$0.2552 \times 10^{-6}$	6.9780	$0.4709 \times 10^{-6}$	6.7048
	0.125	$0.2019 \times 10^{-8}$	6.9812	$0.3964 \times 10^{-8}$	6.8919
4	2	0.0126055	-	0.0238191	-
	1	$0.3002 \times 10^{-4}$	8.7137	$0.3034 \times 10^{-4}$	9.6164
	0.5	$0.6153 \times 10^{-7}$	8.9305	$0.3858 \times 10^{-7}$	9.6192

compare the numerical solutions at the first period near  $x = 0$  with that of the 20th period, noted by dark lines in Figure 6.1. Specifically, we measure the error according to pressure  $p$  as:

$$\begin{aligned}
 E &= \sqrt{\int_0^{\lambda_0} |p_h(x, t) - p_h(x + 20\lambda_0, t)|^2 dx} \\
 (6.4) \quad &= \sqrt{\frac{h}{2} \sum_{n=0}^{n_0-1} \int_{-1}^1 |p_h^n(\xi, t) - p_h^{n+20n_0}(\xi, t)|^2 d\xi}
 \end{aligned}$$

where  $n$  is the element index and  $n_0 = \frac{\lambda_0}{h}$ . Table 6.1 shows the mesh refinement results for  $p = 1$  to 4. Since the local dispersion relation is accurate to order  $2p + 2$ , the global error measure  $E$  defined in (6.4) will decrease at order  $2p + 1$ . This is observed in all the cases.

**6.2. Reflection at grid discontinuity and comparison with eigenfunctions.** In Figures 6.2 to 6.3, we show the propagation of the sine wave (6.3) through a mesh discontinuity. Since the numerical wave reflection properties are dependent on the flux formula used, we will show cases with the exact as well as inexact characteristics flux formulas. This will be indicated by the value of  $\gamma$  used in the computation. A value of  $|\gamma| = 1$  indicates exact characteristics splitting while a value of  $|\gamma| \neq 1$  indicates inexact characteristics flux. In some calculations, a fairly large grid discontinuity has been used. This is to make reflection errors more visible for the purpose of illustration. In all the calculations, a fifth-order ( $p = 4$ ) scheme is used.

**6.2.1. Exact characteristics flux formula  $|\gamma| = 1$ .** In Figures 6.2(a) and (b), a grid discontinuity is introduced at  $x = 50$  with the ratio of grid spacing being 2 and 5 respectively. The exact characteristics-

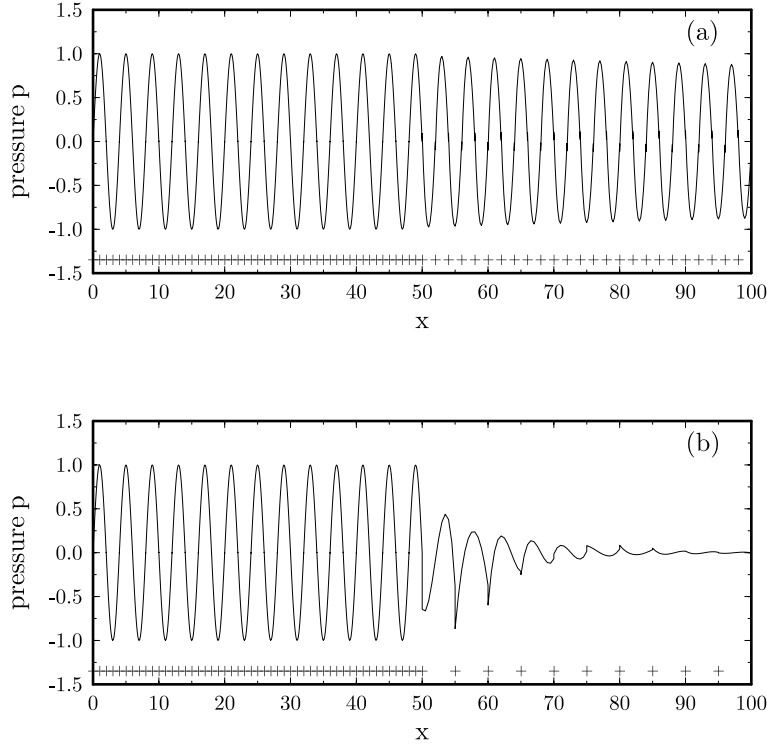


FIG. 6.2. Propagation of a periodic sine wave through a grid discontinuity at  $x = 50$ .  $M = 0$ ,  $\gamma = 1$  (exact characteristics flux).  $+$  indicates grid points. (a)  $h_1 = 1$ ,  $h_2 = 2$ ; (b)  $h_1 = 1$ ,  $h_2 = 5$ .

based flux formula is used in this example with  $\theta = 1$ . In both cases, the abrupt change of element size causes no numerical reflection because the opposite-traveling spurious mode is now non-existent. The damping of wave amplitude in the coarsened grid is due to the reduction in resolution and is expected.

**6.2.2. A slow wave mode  $|\gamma| = 10$ .** In Figure 6.3(a), we show the solution for a case in which  $\gamma$  is large ( $\gamma = 10$ ). This situation is likely to occur when the wave speed of an eigenmode is small relative to the fastest eigenmode governed by a given system of equations. In Figure 6.3(a), the amount of reflection is visible since the grid ratio here is quite large. By subtracting out a calculation with uniform grids (done separately), the reflected wave is extracted and plotted in Figure 6.3(b). Inspecting visually, the reflected wave is in the form of the spurious numerical mode. This will be further confirmed when we compare the numerical solution with the eigenfunction formed in (3.20).

To compare the numerical solution with the eigenfunctions found in section 4, we first extract the complex coefficient vector from the numerical solution by constructing

$$\mathbf{v} = \mathbf{v}|_{t=t_0} + i\mathbf{v}|_{t=t_0 + \frac{T}{4}}$$

in each element. In the above,  $T$  is the period of the sine wave and  $t_0$  is an arbitrary time at which the numerical solution has become time periodic and  $\mathbf{v}$  denotes the solution coefficient vector of the pressure  $p$ . Then, we fit this coefficient vector by a linear combination of the eigenvectors of (3.18). Specifically, suppose the eigenvectors of (3.18) are denoted by  $\mathbf{v}^{(p)}$  and  $\mathbf{v}^{(s)}$  for the physical and spurious modes, we try to find  $a$  and  $b$  such that

$$(6.5) \quad \mathbf{v} = a\mathbf{v}^{(p)} + b\mathbf{v}^{(s)}$$

The coefficients  $a$  and  $b$  are computed by requiring (6.5) be orthogonal to  $\mathbf{v}^{(p)}$  and  $\mathbf{v}^{(s)}$ . In other words, we “decompose” the numerical solution into eigen-modes. This is done for every element and the residues of (6.5) have been found to be near machine zero in all cases. The magnitudes of  $a$  and  $b$  plotted in Figure 6.3(c). Here, circles indicate the magnitude of the physical mode,  $|a|$ , and the triangles the spurious mode,  $|b|$ . The reflection at the interface at  $x = 50$  and their subsequent exponential decay are clearly shown. Also shown, in dotted lines, are the predictions of the reflected and transmitted waves with their amplitudes at the interface being determined by (5.13) and (5.14). Excellent agreements are found.

**6.3. Propagation of an acoustic pulse with mean flow.** In the third example, Figure 6.4, we show the propagation of an acoustic pulse in a mean flow of Mach number  $M = 0.8$ . We solve (6.1)-(6.2) using the Lax-Friedrich formula (2.9) with  $\theta = 1$ . The initial Gaussian profile in the  $u$  velocity component is separated into a downstream propagating pulse, with speed  $M + 1$ , and an upstream propagating pulse, with speed  $M - 1$ . Both pulses are to propagate through a grid discontinuity of ratio  $h_2/h_1 = 5$  located at  $x = 30$  and  $x = -30$  respectively. The difference in wave propagation speed results in two different upwind factors  $\gamma$  for the two pulses, namely,  $\gamma = 1$  for the downstream propagating pulse and  $\gamma = -9$  for the upstream propagating pulse according to (3.15) and (3.16). For the right-traveling pulse, since the flux formula is the exact characteristics splitting, no reflection occurs as the pulse propagates through the grid discontinuity. For the left-traveling pulse, small reflected waves are detected due to the inexact characteristics flux formula for that wave speed. We note that the reflected waves are in the form of spurious waves and decay rapidly. We emphasize that the use of a relatively large abrupt increase in grid size is to make the reflections more visible. Indeed, a calculation using a grid ratio of 2 produced much smaller reflected waves.

**7. Conclusions.** We have carried out a detailed study of spatially propagating waves in a discontinuous Galerkin scheme applied to a system of linear hyperbolic equations. An eigenvalue problem for the spatially propagating waves is formulated. In one dimensional space, the eigenvalue problem reduces to a quadratic equation and, consequently, yields at most two numerical wave modes for each physical wave mode of the partial differential equations. One is physically significant with the dispersion error that decays like  $h^{2p+3}$  and the dissipation error that decays like  $h^{2p+2}$  locally. The other numerical mode is spurious. The spurious mode becomes non-existent when the exact characteristics splitting flux formula is used. Furthermore, reflection and transmission coefficients of an incident wave at an interface of grid discontinuity are derived. It is shown that numerical reflection error consists of only the spurious mode and its magnitude depends on the spatial resolution of the grids on both sides of the interface. Theoretical predictions are verified with numerical examples. These predictions should benefit the design and application of the DGM scheme with non-uniform grids. In a forthcoming paper, we will examine the effects of grid discontinuity in two-dimensional space.

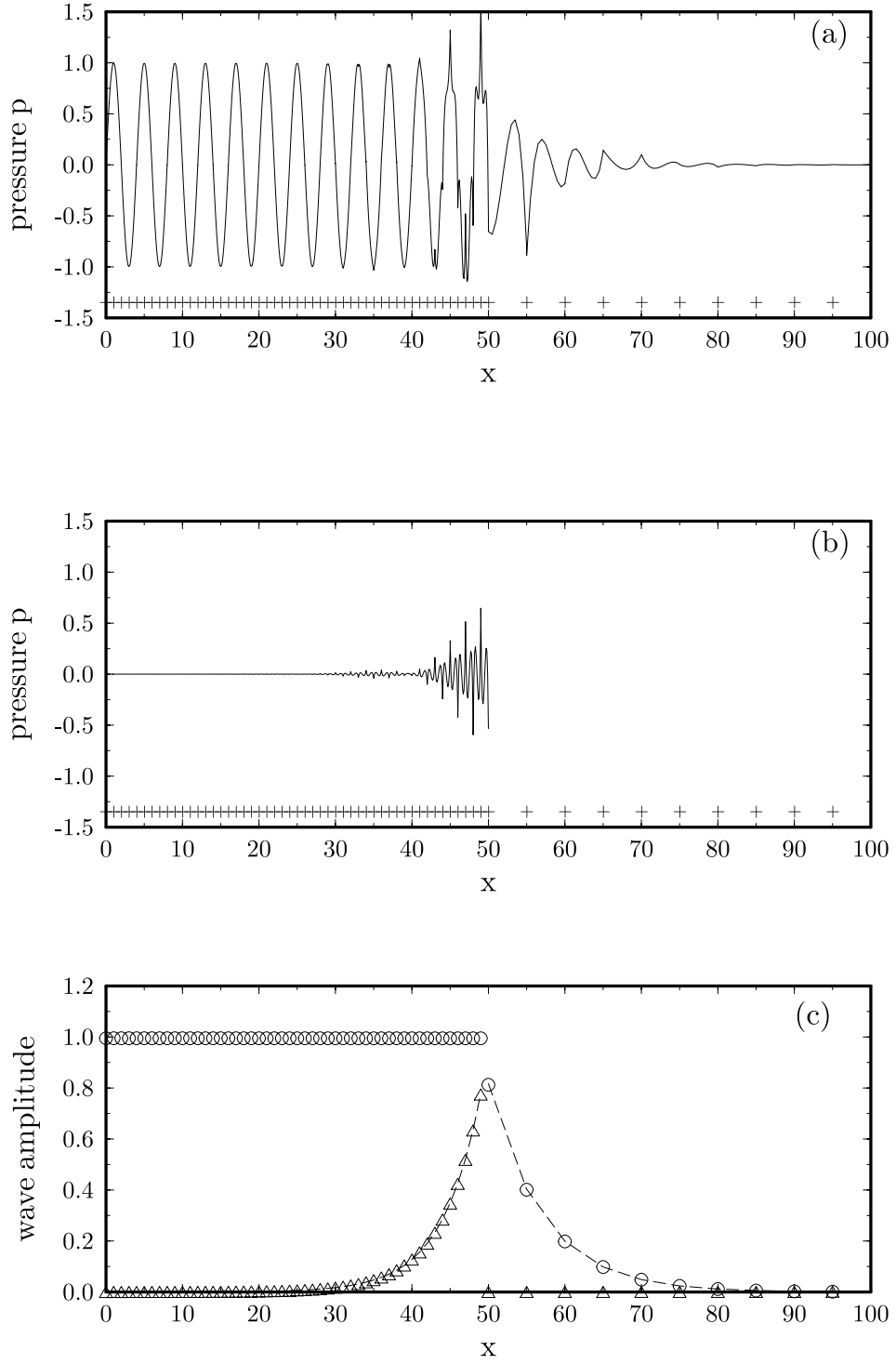


FIG. 6.3. (a) Propagation of a sine wave through a grid discontinuity,  $h_2/h_1 = 5$ ,  $\gamma = 10$ . (b) Reflected wave. (c) Decomposition of numerical solution into physical and spurious modes. circles, physical mode; triangles, spurious mode; dashed lines, theoretical predictions of (5.13)-(5.14).

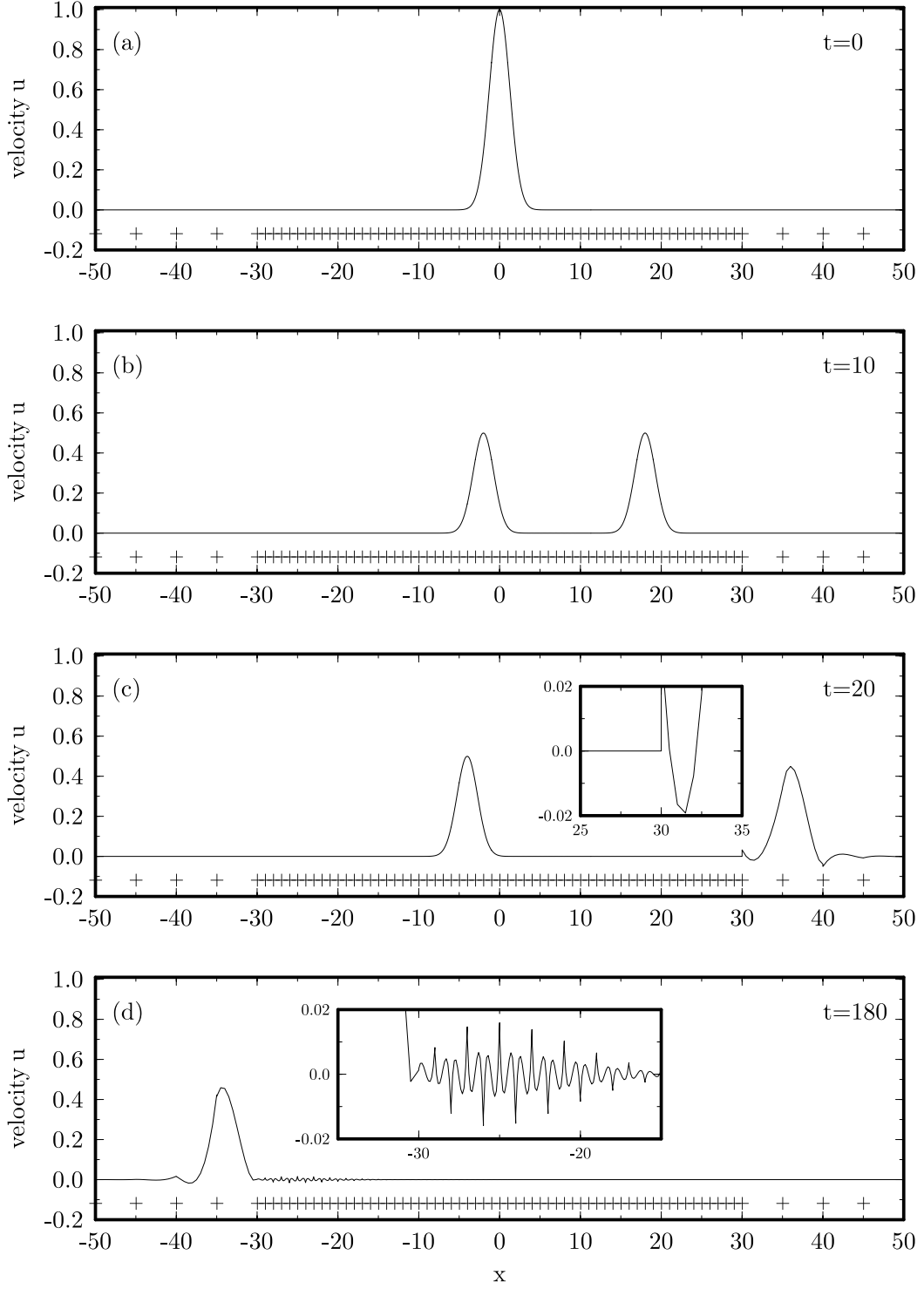


FIG. 6.4. Propagation of an acoustic pulse in a mean flow of  $M = 0.8$ , using Lax-Friedrich flux formula (2.9) with  $\theta = 1$ . Inserts indicate magnified ranges of interest.

## REFERENCES

- [1] S. Adjerid, M. Aiffa and J. E. Flaherty, High-order finite element methods for singularly perturbed elliptic and parabolic problems, *SIAM J. Applied Math.*, 55, No. 2 (1995), pp. 520-543.
- [2] H. Atkins and C.-W. Shu, Quadrature-free implementation of discontinuous Galerkin methods for hyperbolic equations, *AIAA Journal*, 36 (1998), pp. 775-782.
- [3] F. Bassi and S. Rebay, A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier-Stokes equations, *Journal of Computational Physics*, 131 (1997), pp. 267-279.
- [4] R. Biswas, K. D. Devine and J. E. Flaherty, Parallel, adaptive finite element methods for conservation laws, *Applied Numerical Math.*, 14 (1994), pp. 255-283.
- [5] B. Cockburn, G. E. Karniadakis and C.-W. Shu, The development of discontinuous Galerkin methods, *Discontinuous Galerkin Methods* (Cockburn et al. eds), Springer, pp. 3-50, 2000.
- [6] B. Cockburn, M. Luskin, C.-W. Shu and E. Süli, Post-processing of Galerkin methods for hyperbolic problems, *Discontinuous Galerkin Methods* (Cockburn et al. eds), Springer, pp. 291-300, 2000.
- [7] B. Cockburn and C.-W. Shu, The Runge-Kutta discontinuous Galerkin method for the conservation Laws V, *Journal of Computational Physics*, 141 (1998), pp. 199-224.
- [8] P. G. Drazin and W. H. Reed, *Hydrodynamic stability*, Cambridge University Press, 1981.
- [9] F. Q. Hu, M. Y. Hussaini and P. Rasetarinera, An analysis of the discontinuous Galerkin method for wave propagation problems, *Journal of Computational Physics*, 151 (1999), pp. 921-946.
- [10] F. Q. Hu, M. Y. Hussaini and J. L. Manthey, Low dissipation and low dispersion Runge-Kutta schemes for computational acoustics, *Journal of Computational Physics*, 124 (1996), pp. 177-191.
- [11] F. Ihlenburg and I. Babuska, Finite element solution of the Helmholtz equation with high wave number part II: the h-p version of the FEM, *SIAM J. Numer. Anal.*, 34, No. 1 (1997), pp. 315-358.
- [12] C. Johnson and J. Pitkäranta, Convergence of a fully discrete scheme for two-dimensional neutron transport, *SIAM J. Numer. Anal.*, 20 (1983), pp. 951-966.
- [13] P. Lesaint and P. A. Raviart, On a finite element method for solving the neutron transport equation, in *Mathematical Aspect of Finite Elements in Partial Differential Equations*, C. de Boor, ed., Academic Press, New York, pp. 89-123, 1974.
- [14] R. B. Lowrie, P. L. Roe and B. van Leer, A space-time discontinuous Galerkin method for the time-accurate numerical solution of hyperbolic conservation laws, *AIAA paper 95-1658*, 1995.
- [15] R. B. Lowrie, Compact higher-order numerical methods for hyperbolic conservative laws, PhD thesis, University of Michigan, 1996.
- [16] Maple V, Language Reference Manual, Waterloo Maple Inc.
- [17] T. E. Peterson, A note on the convergence of the discontinuous Galerkin method for a scalar hyperbolic equation, *SIAM J. Numer. Anal.*, 38, No. 1 (1991), pp. 133-140.
- [18] P. Rasetarinera, M. Y. Hussaini and F. Q. Hu, Some remarks on the accuracy of a discontinuous Galerkin method, *Discontinuous Galerkin Methods* (Cockburn et al. eds), Springer, pp. 407-412, 2000.
- [19] G. R. Ritcher, An optimal-order error estimate for the discontinuous Galerkin method, *Mathematics of Computation*, 50 (1988), pp. 75-88.
- [20] S. Sherwin, Dispersion analysis of the continuous and discontinuous Galerkin formulations, *Discontinuous Galerkin Methods* (Cockburn et al. eds), Springer, pp. 425-431, 2000.
- [21] L. L. Thompson and P. M. Pinsky, Complex wavenumber Fourier analysis of the p-version finite element

method, Computational Mechanics, 13 (1994), pp. 255-275.

- [22] T. Warburton and G. E. Karniadakis, A discontinuous Galerkin method for the viscous MHD equations, Journal of computational Physics, 152 (1999), pp. 608-641.

## Appendix.

**A1. Kronecker product.** Let  $\mathbf{A} = \{\alpha_{ij}\}_{l \times k}$  and  $\mathbf{B} = \{\beta_{ij}\}_{n \times m}$ . The Kronecker product is defined as

$$(7.1) \quad \mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} \alpha_{11}\mathbf{B} & \alpha_{12}\mathbf{B} & \dots & \alpha_{1k}\mathbf{B} \\ \alpha_{21}\mathbf{B} & \alpha_{22}\mathbf{B} & \dots & \alpha_{2k}\mathbf{B} \\ \dots & \dots & \dots & \dots \\ \alpha_{l1}\mathbf{B} & \alpha_{l2}\mathbf{B} & \dots & \alpha_{lk}\mathbf{B} \end{bmatrix}_{ln \times km}$$

It is easy to verify by direct calculation that for any matrices  $\mathbf{A}$ ,  $\mathbf{B}$  and vectors  $\mathbf{x}$ ,  $\mathbf{y}$ , we have

$$(7.2) \quad (\mathbf{A} \otimes \mathbf{B})(\mathbf{x} \otimes \mathbf{y}) = (\mathbf{A}\mathbf{x}) \otimes (\mathbf{B}\mathbf{y})$$

**A2. Polynomials  $G(x)$  and  $H(x)$ .** Polynomials  $G(x)$  and  $H(x)$  appeared in (4.2):

	$G(x)$	$H(x)$
$p = 1$	$1 - \frac{1}{3}x$	$1 + \frac{2}{3}x + \frac{1}{6}x^2$
$p = 2$	$1 - \frac{2}{5}x + \frac{1}{20}x^2$	$1 + \frac{3}{5}x + \frac{3}{20}x^2 + \frac{1}{60}x^3$
$p = 3$	$1 - \frac{3}{7}x + \frac{1}{14}x^2 - \frac{1}{200}x^3$	$1 + \frac{4}{7}x + \frac{1}{7}x^2 + \frac{2}{105}x^3 + \frac{1}{840}x^4$
$p = 4$	$1 - \frac{4}{9}x + \frac{1}{12}x^2 - \frac{1}{126}x^3 + \frac{1}{3024}x^4$	$1 + \frac{5}{9}x + \frac{5}{36}x^2 + \frac{5}{252}x^3 + \frac{5}{3024}x^4 + \frac{1}{15120}x^5$
$p = 5$	$1 - \frac{5}{11}x + \frac{1}{11}x^2 - \frac{1}{99}x^3 + \frac{1}{1584}x^4 - \frac{1}{55440}x^5$	$1 + \frac{6}{11}x + \frac{3}{22}x^2 + \frac{2}{99}x^3 + \frac{1}{528}x^4 + \frac{1}{9240}x^5 + \frac{1}{332640}x^6$

It is straightforward to verify that  $H(x)/G(x)$  is exactly the Pade approximation of  $e^x$  to order  $2p+2$ . This has been confirmed up to  $p = 16$  and is conjectured to be true for all  $p$ .

**A3. Accuracy of eigenvalues.** In this appendix, we give a derivation of (4.5) and (4.6). We need only to consider the cases when  $|\gamma| \neq 1$ . The case for  $|\gamma| = 1$  follows trivially from (4.2).

For convenience of discussion, define

$$\beta = (-1)^{p+1} \frac{1+\gamma}{1-\gamma}.$$

Then, equation (4.3) becomes

$$(7.3) \quad \lambda^2 - \left[ \frac{H(iK)}{G(iK)} + \beta \frac{H(-iK)}{G(iK)} \right] \lambda + \beta \frac{G(-iK)}{G(iK)} = 0.$$

Let the two roots of (7.3) be denoted by  $\lambda^{(p)}$  and  $\lambda^{(s)}$  with their values at  $K = 0$  as follows,

$$\text{at } K = 0, \quad \lambda^{(p)} = 1 \quad \text{and} \quad \lambda^{(s)} = \beta.$$

Now consider an auxiliary quadratic equation:

$$(7.4) \quad \sigma^2 - \left[ e^{iK} + \beta e^{-iK} \frac{G(-iK)}{G(iK)} \right] \sigma + \beta \frac{G(-iK)}{G(iK)} = 0.$$

The two roots of (7.4) are easily found to be

$$\sigma^{(p)} = e^{iK} \quad \text{and} \quad \sigma^{(s)} = \beta e^{-iK} \frac{G(-iK)}{G(iK)}.$$

By subtracting (7.4) from (7.3), we get

$$(7.5) \quad \lambda^2 - \sigma^2 - \left[ \frac{H(iK)}{G(iK)} + \beta \frac{H(-iK)}{G(iK)} \right] \lambda + \left[ e^{iK} + \beta e^{-iK} \frac{G(-iK)}{G(iK)} \right] \sigma = 0.$$

By the results of Appendix A2, we have

$$\frac{H(iK)}{G(iK)} = e^{iK} + R(iK)$$

and

$$\begin{aligned} \frac{H(-iK)}{G(iK)} &= \frac{H(-iK)}{G(-iK)} \frac{G(-iK)}{G(iK)} = [e^{-iK} + R(-iK)] \frac{G(-iK)}{G(iK)} \\ &= e^{-iK} \frac{G(-iK)}{G(iK)} + R(-iK) \frac{G(-iK)}{G(iK)} \end{aligned}$$

where  $R(x)$  is an  $O(x^{2p+2})$  function with *real* coefficients. Then equation (7.5) can be written as

$$(7.6) \quad \lambda^2 - \sigma^2 - \left[ e^{iK} + \beta e^{-iK} \frac{G(-iK)}{G(iK)} \right] (\lambda - \sigma) = \left[ R(iK) + \beta R(-iK) \frac{G(-iK)}{G(iK)} \right] \lambda.$$

For simplicity, let's define

$$A(iK) = e^{iK} + \beta e^{-iK} \frac{G(-iK)}{G(iK)},$$

$$B(iK) = R(iK) + \beta R(-iK) \frac{G(-iK)}{G(iK)},$$

and rewrite (7.6) as

$$(7.7) \quad (\lambda - \sigma) [\lambda + \sigma - A(iK)] = B(iK) \lambda.$$

It is straightforward to verify that, as  $K \rightarrow 0$ , we have

$$(7.8) \quad A(iK) = [1 + \beta] + \mu_1(iK) + \mu_2(iK)^2 + \dots$$

$$(7.9) \quad B(iK) = [1 + \beta] \nu_1(iK)^{2p+2} + \nu_2(iK)^{2p+3} + \dots$$

and

$$(7.10) \quad \lambda^{(p)} + \sigma^{(p)} - A(iK) = [1 - \beta] + \mu'_1(iK) + \mu'_2(iK)^2 + \dots$$



$$(7.11) \quad \lambda^{(s)} + \sigma^{(s)} - A(iK) = [\beta - 1] + \mu_1''(iK) + \mu_2''(iK)^2 + \dots$$

where all the coefficients on the right hand sides are real and the dots represent higher order terms in  $(iK)$  with real coefficients.

Therefore,

(i) if  $\beta \neq \pm 1$ , it follows easily from (7.7) that

$$(7.12) \quad \lambda^{(p)} - \sigma^{(p)} = \frac{[1 + \beta]}{[1 - \beta]} \nu_1 (iK)^{2p+2} + \dots$$

and

$$(7.13) \quad \lambda^{(s)} - \sigma^{(s)} = \frac{[1 + \beta]}{[\beta - 1]} \beta \nu_1 (iK)^{2p+2} + \dots$$

This immediately leads to (4.5) and (4.6).

(ii) if  $\beta = 1$ , then, instead of (7.10) and (7.11), we have

$$\lambda^{(p)} + \sigma^{(p)} - A(iK) = \mu_1'(iK) + \mu_2'(iK)^2 + \dots$$

$$\lambda^{(s)} + \sigma^{(s)} - A(iK) = \mu_1''(iK) + \mu_2''(iK)^2 + \dots$$

which gives

$$\lambda^{(p)} - \sigma^{(p)} = \frac{2}{\mu_1'} \nu_1 (iK)^{2p+1} + \dots$$

and

$$\lambda^{(s)} - \sigma^{(s)} = \frac{2}{\mu_1''} \beta \nu_1 (iK)^{2p+1} + \dots$$

This is one order lower than that of case (i).

(iii) if  $\beta = -1$ , then, instead of (7.9), we have

$$B(iK) = \nu_2 (iK)^{2p+3} + \dots$$

and it follows from (7.7) that

$$\lambda^{(p)} - \sigma^{(p)} = \frac{1}{2} \nu_2 (iK)^{2p+3} + \dots$$

and

$$\lambda^{(s)} - \sigma^{(s)} = -\frac{1}{2} \nu_2 (iK)^{2p+3} + \dots$$

This is one order higher than that of case (ii).

We note, however, that (ii) or (iii) is possible only if  $\gamma = 0$  or  $\gamma = \infty$ . Both are not very practical situations.